DOCUMENT RESUME

ED 359 856 HE 026 489

AUTHOR	Greenwood, Addison
TITLE	National Assessment of College Student Learning: Getting Started. A Summary of the Beginning Activities.
INSTITUTION	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	ISBN-0-16-041769-4; NCES-93-116
PUB DATE	May 93
NOTE	192p.; For a related document, see ED 346 808.
AVAILABLE FROM	Superintendent of Documents, U.S. Government Printing Office, Mail Stop: SSOP, Washington, DC 20402-9328.
PUB TYPE	Reports - General (140)
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Academic Standards; *College Outcomes Assessment; College Students; Competency Based Education; *Educational Assessment; *Educational Objectives; Educational Planning; Evaluation Methods; Higher Education; Outcomes of Education; Program Development; Skill Development; Standards; *Student Evaluation
IDENTIFIERS	*National Education Goals 1990

ABSTRACT

This report explores the issues and concerns related to the development of a process to assess college student learning. Its primary focus is the attainment of National Education Goal 5.5 by the year 2000 which reads, "The proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially." The primary source of information for this report came from a set of 15 papers commissioned as background for a study design workshop held in November 1991, 45 reviews of the papers, and the proceedings of the study design workshop, "National Assessment of College Student Learning: Issues and Concerns." Chapter 1 addresses what it means to undertake a national assessment of college student learning and raises issues inherent in such a national assessment. Chapter 2 considers what specific skills should be assessed (critical thinking skills, assessment in the workplace, assessment in the colleges--basic skills and general intellectual skills, literacy and writing assessments, and necessary research). Chapter 3 raises six standards and other measurement issues: (1) relationship of standards to the task of defining a national assessment of college student learning; (2) historical context for standards; (3) relationship of standards of National Assessment of College Student Learning to the overall charge of Goal 5; (4) the testing of subject-specific content domains; (5) reasonableness of a single set of standards; (6) and the debate over portfolio assessment and its relationship to standards and values issues. (Contains over 450 footnotes.) (GLR)

ED359856

NATIONAL CENTER FOR EDUCATION STATISTICS

National
Assessment
College Student
Learning
Getting
Started

A Summary of the Beginning Activities

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

HE026 489

U.S. Department of Education
Office of Educational Research and Improvement
NCES 93-116

BEST COPY AVAILABLE
2

National Assessment of College Student Learning:

Getting Started

A Summary of the Beginning Activities

Author: Addison Greenwood

Project Officer: Sal Corrallo
National Center for Education Statistics

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Emerson J. Elliott
Acting Assistant Secretary

National Center for Education Statistics
Emerson J. Elliott
Commissioner

National Center for Education Statistics

"The purpose of the Center shall be to collect, analyze, and disseminate statistics and other data related to education in the United States and in other nations."—Section 406(b) of the General Education Provisions Act, as amended (20 U.S.C. 1221e–1).

May 1993

For sale by the U.S. Government Printing Office
Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328
ISBN 0-16-041769-4

FOREWORD

National Education Goal 5 states that "By the year 2000, every adult American will be literate and possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship. Objective 5 reads, "The proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially." This report is the third product to come from work commissioned by the National Center for Education Statistics in 1991 and 1992, in response to the need to develop a means to assess college student learning in support of National Education Goal 5.5.

The purpose of this report was to provide background information for a study design workshop, the second in a series, held in November 1992. The primary source of information for this publication came from a set of 15 papers commissioned as background for a study design workshop held in November 1991, 45 reviews of the papers, and the proceedings of the study design workshop (National Assessment of College Student Learning: Issues and Concerns, NCES 92-068, GPO, Washington D.C 1991.). This document offers neither a summary of the comments nor draws conclusions. That would be premature as we continue to explore the issues and concerns related to the development of process to assess college student learning. Rather it reflects the wide diversity of thinking expressed in the first workshop.

While the first workshop included questions on both content and methodology, there was agreement that the initial step in the process should be the identification of the higher order thinking and communication skills college graduates need to function from a global and a national perspective. That will be the focus of the second workshop. In support of that agenda, this publication begins with a listing of the comments made on what it means to undertake such a project, moves to a discussion of which skills should be assessed and closes with a chapter on selecting standards and other measurement issues. While specific skills and standards are not identified, a variety of related issues and concerns are discussed in great detail. In addition to workshop participants this material is being published as it will be of value to all concerned with the assessment of college student learning at the institution, state or national levels. Comments on any aspect of the project are welcomed. They may be sent to Sal Corrallo, Project Director NACSL, NCES, 555 New Jersey Ave, NW, Washington, D.C. 20208.

A very special thanks to Addison Greenwood for his hard and creative work in refashioning and collating the materials from 15 papers, 45 reviews, and comments of 80 participants to identify what college graduates should know and how their achievement levels amy be identified. His multi-talented thinking and communication skills will be much appreciated by those seeking enlightenment on this rather complex and important activity.

Emerson J. Elliott
Commissioner of Education Statistics
National Center for Education Statistics

PREFACE

Neil Postman is a longtime observer and an incisive critic of the American education scene. In a recent essay, Postman allies himself with H.G. Wells, "who said that civilization is in a race between education and disaster, and that although education is far behind, it is not yet out of the running."[1] Against the dangers presented to the values of modern culture by the debasement of language, Postman believes "we may mount a practical counteroffensive by better preparing the minds of those for whom such language is intended."[2]

He develops "seven insights or principles. . .that are essential to the workings of the critical intelligence,"[3] by which he seems to mean a sort of thinking that has been elemental to modern history, insights he believes "are in the jurisdiction of every teacher at every level of school."[4] As an import from outside of the debate about the national assessment of college student learning, these insights—originally intended as a sort of Swiftian Modest Proposal about revising all American school curricula—provide a refreshing view of how that debate, including NCES development efforts, might proceed.

(1) "A definition is not a manifestation of nature but merely and always an instrument for helping us to achieve our purposes. I.A. Richards once remarked,

> We want to do something, and a definition is a means of doing it. If we want certain results, then we must use certain definitions. But no definition has any authority apart from a purpose, or any authority to bar us from other purposes.

This is one of the most liberating statements I know. . .What students need to be taught, then is that definitions are not given to use by God; that we may depart from them without risking our immortal souls; that the authority of a definition rests entirely on its usefulness, not on its correctness (whatever that means); and that it is a form of stupidity to accept without reflection someone else's definition of a word, a problem, or a situation."[5]

---

[1] Neil Postman, Conscientious Objections, (New York: Vintage Books, 1988): 21.

[2] Ibid., 21.

[3] Ibid., 25.

[4] Ibid., 21.

> (2) "The form in which we ask our questions will determine the answers we get. To put it more broadly: all the knowledge we ever have is a result of questions. Indeed, it is a commonplace among scientists that they do not see nature as it is, but only through the questions they put to it. I should go further: we do not see *anything* as it is except through questions we put to it. . .As Francis Bacon put it more than 350 years ago: 'There arises from a bad and inapt formation of words a wonderful obstruction of the mind.'

This is as good a definition of stupidity as I know: a bad and inapt formation of words. Let us, then, go "back to Bacon," and study the art of question-asking. But we must also focus on the specific details of asking questions in different subjects. What, for example, are the sorts of questions that obstruct the mind, or free it, in the study of history? How are these questions different from those one might ask of a mathematical proof, or a literary work, or a biological theory? The principles and rules of asking questions obviously differ as we move from one system of knowledge to another, and this ought not to be ignored."[6]

> (3) Which leads me to my third principle: namely, that the most difficult words in any form of discourse are rarely the polysyllabic ones that are hard to spell and which send students to their dictionaries. The troublesome words are those whose meanings appear to be simple, like "true," "false," "fact," "law," "good," and "bad." . . .I think it would be entirely practical to design a curriculum based on an inquiry into, let us say, fifty hard words, beginning with "good" and "bad" and ending with "true" and "false."

. . .Show me a student who knows something about what these words imply, what sources of authority they appeal to, and in what circumstances they are used, and I will show you a student who is an epistomologist—which is to say, a student who knows what textbooks try to conceal. And a student who knows what textbooks try to conceal will know what advertisers try to conceal, and politicians and preachers as well."[7]

---

[5] Ibid, 25.

[6] Ibid, 26 - 27.

[7] Ibid, 27 - 28.

(4) "Fourth, I think it would also be practical to design a curriculum based on inquiry into the use of metaphor. Unless I am sorely mistaken, metaphor is at present rarely approached in school except by English teachers during lessons in poetry. This strikes me as absurd, since I do not see how it is possible for a subject to be understood in the absence of any insight into the metaphors on which it is constructed. All subjects are based on powerful metaphors that direct and organize the way we will do our thinking. . .All forms of discourse are metaphor-laden, and unless our students are aware of how metaphors shape arguments, organize perceptions, and control feelings, their understanding is severely limited.[8]

(5) "Which gets me to my fifth concept, what is called reification. Reification means confusing words with things. It is a thinking error with multiple manifestations, some merely amusing, others extremely dangerous. . .Scholars [may] far too often obscure the emptiness of what they are talking and writing about by affixing alluring names to what is not there. I suggest, therefore, that reification be given a prominent place in our studies, so that students will know how it both works and works them over."[9]

(6) "Sixth, some attention must be given to the style and tone of language. Each universe of discourse has its own special way of addressing its subject matter and its audience. Each subject in a curriculum is a special manner of speaking and writing, with its own rhetoric of knowledge, a characteristic way in which arguments, proofs, speculations, experiments, polemics, even humor, are expressed."[10]

Postman's seventh and final principle has to do with the "non-neutrality of media," by which he means that "the form in which information is coded has, itself, an inescapable bias."[11] Among the bones he is picking with how modern America constitutes its culture is that the technology and the forms by which important messages are transmitted are having an inevitably crucial, and inordinately unconscious, impact on individual Americans. His mission is to bring that unconscious drama out onto the stage for critical examination.

---

[8] Ibid, 29-30.

[9] Ibid, 30-31.

[10] Ibid., 32.

[11] Ibid., 32.

This mission statement also fairly applies to many of the scholars who participated in the NCES workshop and the national assessment of college student learning discussion. Clearly the task is prodigious, and the stakes vital. Not to be cute, there is a problem to be solved, and critical thinking about how to do so was what was being asked of them. Their attempts to communicate effectively their views, and in turn the effectiveness of this document in discussing them, will be judged by the outcome. Ultimately, years hence, when the national assessment of college student learning will have fashioned whatever its impact on American education is to be, the archaeologists will find herein some early footprints, hopefully heading in the right direction.

TABLE OF CONTENTS

Foreword . . . iii

Preface . . . v

Introduction . . . 1

1. What does it mean to undertake a National Assessment of College Student Learning? . . . 9
   - Six who say: "Let's proceed, but watch out for. . ." . . . 10
   - "A Noble Venture, but perhaps premature" . . . 18
   - "Devise More Pragmatic and Immediate Systems" . . . 21
   - "Improving Instruction Should be the Major Premise" . . . 30
   - "Real Life Relevance Should Be the Engine" . . . 37
   - "To Improve Instruction, First You Must Locate It" . . . 39
   - "Effective Writing Cannot Be Examined in a Test" . . . 48
   - "Could Shangri-La, Wisconsin, be Exported to America?" . . . 52

2. Which Skills Should Be Assessed? . . . 67
   - The Critical Thinking Skills . . . 71
     - The Search for Definition and Consensus . . . 71
     - Practicality as a Litmus Test . . . 81
   - Assessment in the Workplace . . . 88
     - The Current State of Workplace Assessment . . . 90
   - Assessment in the Colleges . . . 97
     - Basic Skills First, then General Intellectual Skills . . . 99
     - Possible Models at Individual Schools . . . 102
   - Literacy and Writing Assessments . . . 107
     - Writing Assessment to Probe Effective Communication . . . 112
   - The Call for Necessary Research . . . 114
     - Fitting National Assessment of College Student Learning to American Higher Education . . . 115

3. Standards and Other Measurement Issues: Six Important Questions 121

a. Are standards inherent to the elemental task of defining a national assessment of college student learning? 125

b. What is the historical context for deriving and implementing standards? 133

c. How would standards in a proposed National Assessment of College Student Learning relate to the overall charge of Goal 5—do they "transfer" beyond the academic setting? 137

d. More specifically, must the National Assessment of College Student Learning test be subject-specific content domains in order to generate robust and reliable conclusions about transfer? 146

e. Is a single set of standards reasonable, or even possible, given the diversity of institutions and of those to be assessed? 156
    Imposing Social Judgments about Value 158
    The "Multiple Measures" Dilemma 159
    Clarifying the Contributions of College 162
    The Diversity of Populations 165

f. What does the debate over portfolio assessment reveal about the standards and value issues? 167

4. Summing Up 175

Given these considerations, can an assessment of college student learning proceed? 175

INTRODUCTION

In 1990, the National Education Goals Panel (NEGP) established long term objectives to guide America towards educational excellence. National Educational Goal 5 from *America 2000*:

> By the year 2000, every adult American will be literate and will possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship.

Embraced within Goal 5 are the following five objectives:

> (1) Every major American business will be involved in strengthening the connection between education and work.
>
> (2) All workers will have the same opportunity to acquire the knowledge and skills, from basic to highly technical, needed to adapt to emerging new technologies, work methods, and markets through public and private educational, vocational, workplaces or other programs.
>
> (3) The number of quality programs, including those at libraries, that are designed to serve more effectively the needs of the growing number of part-time and mid-career students will increase substantially.
>
> (4) The proportion of those qualified students, especially minorities, who enter college, who complete at least two years, and who complete their degree programs will increase substantially.
>
> (5) The proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems, will increase substantially.

One of the primary authors for the National Center for Education Statistics (NCES) workshop, Stephen Dunbar, opened his paper with an attempt to put this charge in perspective. In so doing, he framed the task at hand:

> It should come as no surprise to careful observers of the current assessment climate in the United States that eventually the federal government would turn its attention to the matter of assessment and accountability in postsecondary education. Despite the many obstacles that have always faced attempts to measure the impact of a college education, or even to articulate the value of a college education to a general audience, we are now faced with an initiative that poses some extremely difficult questions. Some of these questions go well beyond any experiences educators in the United States have had in the arena of assessment and public policy. Many of them present problems of a technical nature that have never been posed to specialists in measurement theory or practice. All of them require answers if the national assessment of college student learning is to become a component of federal educational policy.[12]

In response to the call from the Goals Panel, the National Center for Education Statistics (NCES) convened a large and diverse group of scholars and educators.[13] Their purpose was to develop strategies that could be used to design an assessment to measure Objective 5. To begin the process, two pairs and thirteen individual authors had been previously selected to develop position papers, which were completed over the summer of 1991 and furnished in advance to about a hundred participants who would attend the conference.

Throughout the process, from an authors' briefing session held before the papers were written, all the way through to written comments offered after the conference was held, NCES experienced a dilemma. On the one hand, scholars were chosen for the expertise and viewpoints they had developed and championed during their varied careers, and it was both expected and hoped that their papers could harness that diversity to the task at hand without any sacrifice to the integrity of their many views. On the other hand—in order to corral the sheer volume of ideas necessarily generated by this intellectual tournament—it was felt that certain basic questions considered essential to proceeding with the pragmatic task of instrument development would have to be addressed.

---

[12] Dunbar, Stephen B. "On the Development of a National Assessment of College Student Learning: Measurement Policy and Practice in Perspective," (paper commissioned for the National Assessment of College Student Learning study design workshop by the U.S. Department of Education, National Center for Education Statistics, 1991), p. 1.

[13] In Crystal City, Virginia, November 16-18, 1991. The full list of participants may be found in Appendix.

Thus a structure emerged which imposed something of a generic format onto the analyses, but not an arbitrary one. A synthesis document was written to compress the essential recommendations from the authors' original papers (approximately 500 pages in the aggregate) down to about 100 pages. The outline of the synthesis was also used to organize discussion during the conference, when participants broke up into small group working sessions. The conference proceedings also reflect this format, with separate reports from four working groups as well as summaries by NCES associate commissioners who chaired these groups. First the fifteen primary authors/teams, then three reviewers per paper, and finally the workshop participants considered five aspects of the subject:

(1) Which skills and abilities should be targeted?

(2) What standards should be imposed upon the process?

(3) At what point in their careers should college students be tested?

(4) How to motivate all participants to essay their best performance?

(5) What sort of instrument should be designed to capture these elements?

As the working sessions during the conference unfolded, however, a phenomenon surfaced that had first become evident in trying to capture for the synthesis all relevant points made in the original authors' papers. Scholars, whatever manifest good will and desire to cooperate they may possess, will nonetheless think hard and honestly about these issues. They will do so for any of a number of reasons, but perhaps foremost because many of them have spent their careers grappling with the complexities that underlie such questions. At the workshop were practitioners of many stripes: scholars and theorists who have been involved in conceptualizing critical thinking; professional educators working in state and institution-based education organizations; people at the interface in both government and industry where programs are actually implemented; and professionals who study testing and assessments.

In sum, a group of people possessing an enormous range of experience and many distinct views of American education. The hard-won wisdom from that experience invariably (and fortunately!) found its way into the NCES deliberations, both in the original papers, and throughout the conference. Thus most of the ideas that survived the hurly-burly of debate and argument share a number of features. They tend to be pragmatic in the strongest sense of that term: they are the ways that professionals who embody hundreds of years of collective experience as educators believe that this project can be successful. They do not, however, fit neatly into the outline used as an organizing format.

These many different views about assessing college student learning produce a complex, heterogeneous mosaic, which this document tries to portray. The outline questions for many did provide a rational, useful structure to convey both their recommendations and their caveats. For others, the NCES assignment was seized as an opportunity to address the widest possible audience on the very legitimacy, and direction, of a national assessment of college student learning. Opportunity is perhaps the wrong term. Rather—in the forcefully argued and fervently advocated plans for a national assessment of college student learning proposed by the more radical among the dissenting authors—there is apparent a special sense of duty, a professional mandate to try to reframe the focus they perceived as developing about an assessment of college student learning.

An image that found favor during the workshop was that of a train, stoked up, whistle steaming, ready to leave the station. The question seemed to become: Was it too late to postpone the departure? Many felt it was—that, fueled by the governors, the president and the NAGP, the train of a national assessment was leaving the station, and that their recommendations and advice must be directed to steering it onto the proper track, anticipating difficulties, and trying to smooth the journey. Others, however, argued forcefully that we were not ready, had not clarified where we were heading, and even if we did understand the scope of the journey—did not know how to get there without significant efforts at development and research.

Clearly the debate is not so neatly divided into two camps, in fact at least four factions could be discerned by an observer trying to find some frame with which to capture the rich diversity of the papers and the workshop they spawned. These four more or less distinct approaches to the national assessment of college student learning were not taken up as such during the workshop, but Chapter 1 reveals them.

First, for those who are ready to board the train, and want only to be certain that it goes in the right direction. Many of the critical thinking experts belong to this group. Second, to stay with the image, those who want to delay its departure until we can be more confident of a safe and successful journey. Dunbar epitomized those concerned that such a prodigious undertaking requires more and better thinking and planning and developmental research than has presently been focused on the problem. Third, those who think they can get where we should be going by other modes of transportation, seeking either new or extant sources of data to respond to Goal 5. (The two author/teams, Peter Ewell and Dennis Jones, and Daniel Resnick and Natalie Peterson, provided such concrete, but alternative, ways of travel.) And finally the dissenters, many of whom believed fervently in the promise and potential of a national effort at assessment, but who feel with equal conviction that a monolithic instrument will almost certainly derail the truly valuable momentum towards educational improvement and reform that has been generated by the call for a national assessment of college student learning.

It is premature, however, to think of this as a set of operating instructions.[14] Rather, as it is not at all settled yet what sort of instrument(s) might be built, this paper is better styled as a collection of insights about how to *approach* the design and construction of such an instrument(s). Because no unequivocal and fully funded effort has yet been undertaken to create such an assessment structure, it is wholly appropriate that those calling for a re-examination of the entire undertaking be heard. What could be perceived as equivocation, they suggest, amounts to much more than nay-saying. In fact, it provides the best intellectual challenge that those developing a national assessment of college student learning could face, and in so doing can only serve to clarify the critical thinking that must be brought to the process, if such instrument(s) are eventually to be devised and successfully implemented. Or, their alternative visions may, in the process of advocacy, begin to coalesce into a truly alternative approach to assessing college student learning.

A national assessment of the scope that is taking shape presents an unprecedented opportunity. While there have been many educational reforms envisioned and mounted over the years, the millions of dollars that could be provided for this effort—which has already been certified by significant political commitments—suggest the potential which energized the national assessment of college student learning debate. Participants in the NCES developmental workshop manifestly saw their efforts as a vital step in forming what will one day likely become a prominent and visible feature of American higher education. Regardless of how seriously they wanted to shake up the planning process—some to slow down the building momentum toward developing an instrument—virtually all participants joined together to earnestly and critically contend over how best to proceed.

---

[14] The NCES approach to the workshop encouraged and invited this diversity. However, if an instrument is going to be developed, almost all agreed that considerable work needs to be done. To further provide a useful framework for such an effort, a second developmental workshop is planned for November of 1992. This document thus stands at the crossroads between these two conversations. It can probably only suggest and not fully capture the rich diversity of the first meeting. But it should provide for all those interested in assessment in American education - and especially those who will participate in the second meeting - a summary of the key concerns for those wishing to develop a means of assessing college student learning.

* * *

So the current document presents a snapshot in time (ca. November, 1991) of a wide range of thinking about the national assessment. The original format that was suggested to the authors and imposed onto the workshop discussions has given way to a modification, which is embodied here. The assumption that scholars could more or less directly approach the outline questions turned out to be naive, in one particular respect: far too much consensus, prerequisite to such as structure, was assumed. The papers and the workshop they spawned provide eloquent evidence that no such consensus exists. As described above, people were all over the map, many in fact off of the edge. When they convened in Washington, it became apparent that the premises from which they could proceed to consider the details of a national assessment of college student learning were not shared. To clarify these premises, in effect to frame the national assessment of college student learning, many scholars asked the most probing, difficult and essential questions about the undertaking.

Thus, Chapter 1 provides an irreducible introduction, echoing the question which will have an impact on virtually all the details and subsequent plans in the process: What Does it Mean to Undertake a National Assessment of College Student Learning? In addressing this question, some authors and participants found the questions of the basic format to be vital, others irrelevant, and some contradictory to their views. In effect, they asked and answered their own questions in addressing the meta-question that is the subject of Chapter 1.

In trying to coalesce the many diverse ideas about a national assessment of college student learning and to proceed to specifics, this first chapter needed to written and comprehended, debated and revised, by one and all. At the workshop, this process began in working sessions, and it arrives at a crystallization here. Once the larger vision of a national assessment of college student learning has been sketched out, and the strongly felt, alternative factions have expressed their reservations, caveats, and objections, planners can in some coherent and rational way proceed to specifics. As it developed, the outline was a good first stab at a structure, but the workshop clarified that first things must come first, and for a coherent discussion to proceed, the very first thing is the question of Meaning, as it unfolds in Chapter 1. With this as a vital context, scholars can proceed to the next, more specific steps of trying to specify the structure and content of the assessment.

Chapters 2 and 3 begin this process. Why Skills and Standards? Why not the questions concerning student motivation to perform, when the test should be administered, and what specific test structure should be used? While these latter questions were part of the charge, and were discussed at some length during the workshop, a lesson was learned about first things. Until you can agree on exactly what skills are to be assessed—what constitutes the Objective 5 abilities—you cannot begin to realistically address these other questions. And while this argument might seem even to apply to the question of *Standards*, a careful reading of the papers and the workshop proceedings shows otherwise. Standards, it seems, are fundamental, basic, inherent, and imbedded in the choice of which skills to target. The question of standards seems to be a fundamental and primary lens through which early catalogs of skills must be evaluated. Thus the insights about the more practical, implementation questions have been collected for future consideration. For now, the hard but exciting task of getting one's hands in the clay and beginning to create a structure with content will proceed. Thus the three chapters of this study capture the salient material developed for NCES during more than a year of dialogue through papers, reviews and discussion about:

(1) What is the Meaning of National Assessment of College Student Learning?
    - the basic approaches that seem to be taken by scholars in first confronting the prospect of a National Assessment of College Student Learning.

(2) What are the Objective 5 abilities?
    - which skills and abilities should be targeted?

(3) How should the assessment of the identified abilities be approached, with what standards?
    - in trying to develop standards by which the test will be constructed and scored, what important issues arise about context, content, proficiency levels and transfer?

1. WHAT DOES IT MEAN TO UNDERTAKE A NATIONAL ASSESSMENT OF COLLEGE STUDENT LEARNING ?

The NCES approach to a national assessment of college student learning began with the premise that a project of this magnitude goes far beyond the technicalities of measurement. Thus, from the beginning, scholars were recruited who could develop a dialogue that would be inclusive.[15] The plan was to circle widely around the actual structure of the instrument, trying to flush into the open some of the fundamental, underlying, social and intellectual issues that such an enormous event was likely to unleash.

Thus, while the authors' meeting in August essayed a consensus about how the fifteen author/team(s) would proceed, and the general outline of the five questions was ostensibly approved, storm warnings during that meeting were prominent. Before they had really begun to struggle (in writing) with their positions and their response to the NCES charge, the authors were polite, if meticulous; camaraderie overrode the occasional flare-up over what, in a friendly and preliminary meeting, could be seen as "semantic" distinctions.

This genial atmosphere of collegiality, it became apparent only several months later, actually masked passionate and intensely held convictions about the importance of a national assessment of college student learning, which most of the authors view as an event of historic magnitude in American education. If it can be done, many said, please be sure we do it right, for such a bold and expensive experiment—should it "fail"—could set back progressive education in America. For others, the effort was either premature, wrong-headed, ill-framed, or otherwise misbegotten, and these dissenters expressed their concerns about the fallout from a "failure" in forcefully argued alternative visions of how the federal government should proceed, often with the loftiest educational goals invoked as a rationale.

---

[15] The structural plan sponsored by NCES was unusually substantive, and by the time the primary authors settled in to write the papers that were to become the foundation for the workshop, extensive discussion of their mission had already taken place. Conferring by phone and letter with one another and Washington, they arrived in August for a preparatory meeting. NCES project director Sal Corrallo had organized the group, and provided them with extensive background material. The group meeting was to specifically consider their "marching orders" - a briefing paper that framed the task as NCES saw it. [The collection of their finished papers would be disseminated first to a set of three reviewers per paper for critical comment, and then the papers and reviews to be mailed to the hundred-plus conferees well in advance of the November workshop.] This frame was drawn so as to be wide and flexible, given the impressive background and collective experience of the commissioned authors.

Six who say: "Let's proceed, but watch out for . . ."

Richard Paul has been one of the leading lights during the 1980s, as critical thinking has moved onto the main stage of progressive American education.[16] As Director of the Center for Critical Thinking at Sonoma State University (CA), he and Assistant Director ***Gerald M. Nosich***[17] provided a very rich exposition and analysis of critical thinking, including the fundamental components generally agreed to comprise the various aspects of the discipline. Their catalog begins with 21 objectives they believe any national assessment of college student learning should embrace, and they go on to demonstrate, one by one, how a "*rich, substantive concept of critical thinking*" can meet these criteria.

These descriptors are not mere adjectives, they stress. As warriors of a sort—who have endured the struggle to achieve academic respectability and admission to the halls of the Academy for critical thinking—they emphasize that critical thinking is not merely a catch-all category, but rather a concept with substance. The substance is founded on a solid body of research, which provides educators and analysts something of a perch from which to view lesser, non-substantive versions of critical thinking which would not serve the process of assessing college student learning very well. To illustrate the distinction, the authors warn against certain unfortunate, but predictable, presuppositions:

---

[16] See, for example: Paul, Richard. (1990) Critical Thinking: What Every Person Needs To Survive In A Rapidly Changing World. Rohnert Park, CA: Center For Critical Thinking and Moral Critique.

Paul, Richard, et al. (1990) Critical Thinking Handbook: K-3rd Grades. A Guide for Remodelling Lesson Plans in Language Arts, Social Studies & Science. Rohnert Park, CA: Center For Critical Thinking and Moral Critique.

Paul, Richard, et al. (1990) Critical Thinking Handbook: 4th -6th Grades. A Guide for Remodelling Lesson Plans in Language Arts, Social Studies & Science. Rohnert Park, CA: Center For Critical Thinking and Moral Critique.

Paul, Richard, et al. (1990) Critical Thinking Handbook: 6th-9th Grades. A Guide for Remodelling Lesson Plans in Language Arts, Social Studies & Science. Rohnert Park, CA: Center For Critical Thinking and Moral Critique.

Paul, Richard, et al. (1989) Critical Thinking Handbook: High School A Guide for Redesigning Instruction Rohnert Park, CA: Center For Critical Thinking and Moral Critique.

[17] Also an author in the field; see Nosich, Gerald, *Reasons and Arguments*. (Belmont, CA: Wadsworth, 1981). Their NCES contribution: Paul, Richard W. and Gerald M. Nosich, "A Proposal for the National Assessment of Higher-Order Thinking at the Community College, College, and University Levels (paper commissioned for the NACSL study design workshop by the U.S. Department of Education, National Center for Education Statistics, 1991): 3-9.

- that the meaning or terminology of critical thinking is intuitively obvious (hence not in need of scholarly analysis), or
- that each concept underlying critical thinking (such as assumption, inference, implication, reasoning . . .) can be analyzed separately from a theory that accounts for the interrelation of these concepts, or
- that the skills of critical thinking can be adequately cultivated without reference to the values, traits of mind, and dispositions that underlie those skills.[18]

Their experience and a number of studies suggest that "theoretically superficial" concepts of critical thinking can lead to at least three serious problems:

(1) Important critical thinking concepts, which must be clearly defined to be used effectively in assessment, may be used vaguely, inconsistently, incorrectly, or misleadingly,

(2) a false, misleading, or simplistic over-arching concept of critical thinking may be fostered, and/or

(3) an unrealistic strategy for the assessment and cultivation of critical thinking may be incorporated into testing and teaching.[19]

Paul and Nosich in their paper have provided the National Assessment of College Student Learning process with a rich template to begin actually to specify elements of the assessment. Their analyses of the four component domains of critical thinking, the skills and abilities that emerge from them, and the many questions about standards, domain, and transfer are taken up in detail in Chapters 2 and 3. Much of their analysis also involves consideration of some of the technical issues of measurement and the dilemma of student motivation that have been postponed for later consideration. But they also argue that their proposed assessment strategy would provide a number of levers to improve instruction at the postsecondary level.[20]

---

[18] Ibid., 9.

[19] Ibid., 9.

Don Rock provides a view of the assessment process from deep inside the world of instrument construction and analysis, at the Educational Testing Service (ETS). Here more than anywhere are national level assessment developers likely to find an approach that considers the creation of an assessment instrument from the inside out. His overarching recommendation is that the skills and abilities "should be defined from a teaching/learning perspective, so that their enhancement can be factored into classroom experiences."[21]

Rock concedes that tying assessment to instruction "would seem at 'first blush' to go well beyond the intent of the past national assessments of cognitive skills, e.g., NAEP, NELS, and NALS." However, he notes that "the 1992 NAEP assessment is in some content areas moving closer in the direction of merging assessment and instruction."[22] In a number of ways—for example "continually increasing the proportion of extended constructed (free) response items,"[23]—NAEP seems to be forging "the next step in the evolutionary development of large scale assessments," which Rock identifies as "the use of scoring protocols that are developed specifically to provide diagnostic information for instruction."[24] However, the greater complexity of this sort of test poses issues of scoring, expense, and student motivation, which Rock considers in his analysis. He considers the related issues of standards in some depth, while addressing other aspects of National Assessment of College Student Learning that do not fit neatly into the outline, but he does provide a fairly specific plan:

> It is this writer's suggestion that NCES consider a two or three phase approach. The first phase would be to take advantage of present data bases such as the GRE and NALS with the suggested modifications outlined above. The next phase would include

---

[20] Ibid., see 26-27.

[21] Rock, Donald A., "Development of a Process To Assess Higher Order Thinking Skills for College Graduates," (paper commissioned for the NACSL study design workshop by the U.S. Department of Education, National Center for Education Statistics, 1991): 1.

[22] Ibid., 2.

[23] Ibid., 4.

[24] Ibid., 4.

> designing a NAEP-like study to get an aggregate picture of college level performance on the skills deemed to be important by representatives of both industry and the educational community.

But he also argues that this proposed assessment strategy would provide a number of levers to improve instruction at the postsecondary level.[25]

> This type of design would allow the gathering (at the aggregate level) the maximum amount of information for the least burden. It will not, of course, be optimal for individual level process-outcome type of analysis, but some process-outcome relationships can be estimated for various aggregations. The spiralled design could also serve as a check on the validity of the conclusions based on the "stop-gap" use of readily available data such as GRE & NALS.
>
> A third phase might include the development of a computer assisted adaptive test (CAT) battery which in theory[26] would allow one to get accurate individual performance estimates for a minimum amount of testing time across a relatively broad set of skills at the individuals convenience.
>
> Colleges for the most part have all the necessary hardware to carry out such an endeavor if the software is furnished. Since an individual could be assessed at his or her own convenience at any one of a number of available terminals, cooperation and possibly motivation will be increased. Score reporting would be almost instantaneous since the responses are scored immediately on location and can then be sent to a central location. The necessary item parameters for building the CAT battery could be gotten from the second phase.
>
> The drawback of such a system would be that present technology would force us to rely heavily on the multiple choice framework. Certainly the multiple choice part of the assessment could be carried out this way. The free response part of the assessment could also be carried out on the computer and scored later."[27]

---

[25] Ibid., see 26-27.

[26] See F.M. Lord, *Applications of item response theory to practical testing problems, (Hillsdale, NJ:* Lawrence Erlbaum Associates, Inc, 1980).

[27] Rock, op. cit., 22.

Charles Lenth works in Denver with the State Higher Education Executive Officers Organization (SHEEOO), and provided the assessment debate with a rich context in his survey of recent assessment efforts in the various states. That survey reveals that most states are already—to some degree—in the process of exploring postsecondary assessment. With this as a given, he recommends that the proposal NCES develops should embrace the collective state experience, and should dovetail with present and future state-based needs, activities and personnel. Lenth's major thrust is that higher order intellectual skills are not being assessed effectively by the states, and thus a national assessment of college student learning that provided them useful information in this realm is technically feasible, pragmatically desirable, though of course politically challenging.

Lenth makes a strong case for developing a national assessment of college student learning in close collaboration with those educators, administrators, and planners already involved in state systems. In advising that the federal approach be designed to dovetail with their efforts—and thereby avoid encroaching on their hegemony—he is led to conclude that the focus should be on indicators or outcomes measures of general intellectual skills, not subject-specific domains. Norman Frederiksen agrees: "In my opinion, activities related to the National Goals for Education should not be in the hands of authorities in Washington, nor should results be reported only at the national level. A relationship similar to that of NAEP and NAGB would be more reasonable, and there is no reason not to report results at the state level."[28] Only a few others supported Lenth in this conclusion that the national assessment of college student learning ought to be framed generically, beyond the disciplines, but his pragmatic arguments lent much force to his recommendations. His strongest emphasis, however, was on the standards that he believes should prevail, and thus is his a prominent voice in Chapter 3.

For most of the 1980s ***Ed Morante*** was involved, as Director, with two major assessments conducted in the public postsecondary system in New Jersey. His analysis of this relevant history is much more than a recitation of events, as he recounts the experiences both from the point of view of participant, as well as someone looking for lessons to inform the national assessment of college student learning: "Throughout the paper," he says, "questions and problems are raised and then addressed both conceptually and in terms of practical solutions."[29] The value of such an actual "experiment" in large-scale assessment is incalculable, but remains for national assessment of college student learning developers to mine and assay. Of particular use might be the many reports included in his bibliography, where not only the conclusions but also the deliberations of bodies appointed to consider, develop, and implement the two assessments have been reported in some detail.[30]

---

[28] Frederiksen on Dunbar.

Morante's many relevant conclusions are taken up throughout the discussion where they apply, but his paper's recurrent themes cast an important light on two of the central dilemmas already identified in the national assessment of college student learning process. First, how to deal with the human elements, both the motivation of students and the possible resistance of faculty and administrators. Second, the question of how to actually construct questions that go beyond the multiple choice format, yet which can effectively elicit critical thinking performance without excessive reliance on subject-specific knowledge from a particular content domain. Clearly, "the very existence of these statewide tests is the best answer for whether such testing is feasible," he claims. The more problematic questions are "whether the educators have the will to test and whether the nation can face the test results."[31]

---

[29] Edward A. Morante, "General Intellectual Skills (GIS) Assessment in New Jersey," (paper commissioned for the NACSL study design workshop by the U.S. Department of Education, National Center for Education Statistics, 1991): i.

[30] Morante, Edward A., "General Intellectual Skills (GIS) Assessment in New Jersey," (paper commissioned for the NACSL study design workshop by the U.S. Department of Education, National Center for Education Statistics, 1991): 32. See also:

Board of Higher Education, *A Resolution Establishing a Comprehensive Statewide Assessment Program*, (Trenton, NJ: Department of Higher Education, 1985).

College Outcomes Evaluation Program (COEP) Advisory Committee, *Report to the Board of Higher Education*, (Trenton, NJ: Department of Higher Education, 1987).

College Outcomes Evaluation Program (COEP) Advisory Committee, *Appendices to the Report to the Board of Higher Education*. (Trenton, NJ: Department of Higher Education, 1987).

College Outcomes Evaluation Program (COEP) Council, *Report to the Board of Higher Education on the First Administration of the General Intellectual Skills (GIS) Assessment*, (Trenton, NJ: Department of Higher Education, 1990).

Educational Testing Services, *A Report on the Development of the General Intellectual Skills Assessment*, (Princeton, NJ: Educational Testing Services, 1989).

Morante, Edward A., "The Effectiveness of Developmental Programs: A Two-Year Follow-up Study," *Journal of Developmental Education*, 3(1986): 14-15.

New Jersey Basic Skills Council, *Report to the Board of Higher Education on the Results of the New Jersey College Basic Skills Placement Testing*, Fall 1990 Entering Freshmen; (Trenton, NJ: Department of Higher Education, 1991).

Richard Larson, however, is "not convinced that Morante's perception of the possibility of moving from statewide to nationwide assessment is well supported; the procedures employed in New Jersey seem to have been notably intricate, and the vast diversity of students and political/educational climates across the country might well, as Lenth says, present enormous obstacles to using one state's model nationwide."[32] Such skepticism poses a real dilemma for the national assessment of college student learning, because no other model or future pilot testing program is likely to provide a much greater body of actual experience. It also focuses the spotlight on the reporting of results and the larger question of how the assessment ultimately will be used. New Jersey is not—demographically speaking—a small or simple model, and [following Lenth] if the national assessment of college student learning proposes to the fifty states an assessment system that should be implemented, each could presumably do so facing the same sorts, and size, of state-wide problems encountered in New Jersey.

Peter Cappelli is the Co-Director of the National Center on the Educational Quality of the Workforce. His insights[33] thus comprise an important part of the national assessment of college student learning dialogue, because he explicitly refers to changes in America's global competitiveness. While his paper methodically and comprehensively describes how American industry analyzes jobs and tests employees, it also begins to sketch out the crucial link between education and preparation for the workplace. He anticipates a number of commentators with his general feeling that much needs to be done, but he does not explicitly oppose national assessment of college student learning as a first step.

Several of his conclusions, however, do point to other possible avenues of approach. Primarily, his analysis reveals what many already know: that courses as presently structured in most colleges are not designed around, nor do they stress, job relevant skills. He doesn't believe that such skills are *not* developed in college, but rather that the reporting, analysis, and grading system(s) are not constructed to reveal and quantify them. A reorganization of how information is assembled, Cappelli believes, could be a first step towards improving this communication. He has in mind a portfolio system [anticipating Ed White's strong recommendations] which could provide the sorts of material more useful to the bio-data analyses favored by innovative workplace analysts. Ultimately, the system would begin to feed back and "to encourage the process of education in the classroom to be conducted in ways that develop job-related skills. These steps do not require fundamental changes in the content of courses as much as they do in pedagogy."[34]

---

[31] Morante, 1991, op. cit., i.

[32] Larson on Morante.

[33] Cappelli, Peter, "Assessing College Education: What Can Be Learned from Practices in Industry?" (paper commissioned for the NACSL study design workshop by the U.S. Department of Education, National Center for Education Statistics, 1991).

Richard Venesky is an expert on literacy[35] who looks deep within the historical and procedural experience of literacy assessment for lessons about the national assessment of college student learning, on the assumption that literacy itself will be a testable component and should be accomplished in a certain way. His distinctions between functional and cognitive approaches to defining and probing literacy, as well as his analysis of how the texts that must be used might require subject-specific knowledge, are covered in Chapters 2 and 3. Like Rock, he sees some interesting insights available from the NAEP experience, in particular the Young Adult Literacy Survey of 1985 and its successors.[36]

---

[34] Ibid., 15.

[35] See, for example: Venezky, R.L., C.F. Kaestle and A.M. Sum, *The subtle danger*, (Princeton, NJ: Educational Testing Service, 1987).

Venezky, R.L., D. Wagner and B. Cliberti, *Toward defining literacy.* (Newark, DE: IRA, 1990).

Venezky, R.L. "Catching up and filling in," in Literacy learning after high school. In eds. J. Flood, J. Jensen, D. Lapp, & J.E. Squire, *Handbook of research on teaching the English language arts.* (New York: Macmillan, 1991).

[36] See, for example: Kirsch, I.S., and A. Jungeblut, *Literacy: Profiles of America's young adults--final report*, Report No. 16-PL-0 1. (Princeton, NJ: National Assessment of Educational Progress, 1986).

Kirsch, I.S., and P.B. Mosenthal, P.B. *Understanding document literacy: Variables underlying the performance of young adults*, Research Report No. RR-88-62, (Princeton, NJ: Educational Testing Service, 1988).

Kirsch, I.S., and P.B. Mosenthal, "Exploring document literacy: Variables underlying the performance of young adults," *Reading Research Quarterly, 25(1) (1990): 5-30.*

"A noble venture, but perhaps premature"

Stephen Dunbar as a scholar and historian has a great appreciation of "the many obstacles that have always faced attempts to measure the impact of a college education, or even to articulate the value of a college education to a general audience" in the American past. He tries to emphasize that the national assessment of college student learning "initiative poses some extremely difficult questions . . . [which] go well beyond any experiences educators in the United States have had in the arena of assessment and public policy."[37] The skepticism and arguments against developing a national assessment of college student learning are real, he understands, and he advises the Department to weigh them seriously. Only by facing the implications of these very real structural problems does he feel the process can go forward effectively and with some hope of a permanent impact.

Most of these structural problems derive from a fundamental paradox, explains Dunbar, in the effort to measure critical thinking. He does believe that, *if* proper content domains are delineated, and *if* a genuine consensus can be reached on what the Objective 5 abilities consist of, then social scientists have the tools to develop tables of specifications, to write performance tasks and questions, and thereby to create an instrument. He calls this the utility scale, and thinks it can be accomplished. But he reminds us that Goal 5 does not limit itself so neatly to such an academic terrain, but rather (more ambitiously) calls for the sorts of improvements and gains in effectiveness that amount to a larger scale, the social utility scale.[38] It is not merely a question of establishing standards, but whether (having done so) such standards can be defended from the attacks of critics who will refute the premise that the standards rigorously led to the achievement of whatever the avowed goal or purpose.[39] He does not believe

---

[37] Dunbar, op. cit., 1. He presents "Several arguments against any federally funded, large-scale, census approach to the assessment of college student learning in the United States in order to clarify challenges for the developmental effort. Although these arguments may not always appear to reflect principles of educational measurement *per se*, they have their origins in such principles and in the expected quality and utility of the information that might derive from a census approach for measuring college student learning." [1].

[38] Ibid., 6. "At one time, there was some agreement on methods that might reasonably be used to develop psychological scales for judgments of value," he reports, citing Thurstone, L.L. *The Measurement of Values*, (Chicago: University of Chicago Press, 1959). "But the social landscape in the United States has changed dramatically since that time. It is no longer the case that a self-respecting measurement psychologist would presume to possess a storehouse of procedures that could be used to map an achievement scale onto a universal scale that could withstand rightful challenges to it in a diverse society such as ours." [6]

> that national assessment of college student learning should be conceived in this way. On the contrary, [his] main purpose . . . is to show that the goal statement is unreasonable as a charge for development of national assessment of college student learning, precisely because it establishes a utility scale which is beyond the reach of any technical procedures for standard setting. If national assessment of college student learning is itself expected to support inferences of the kind reflected by the goal statement, it will not be beyond reproach.[40]

Dunbar goes into the intricacies of standards and scales at some length, and his arguments are discussed in Chapter 3. In proposing specific research to address some of the problematic issues inherent to the paradox posed by the call for a national assessment of college student learning, he concedes that

> recommendations for further research aren't likely to satisfy those charged with carrying out a policy initiative at the federal level. What is being recommended here is not basic research of the sort that might lead to implications for a national assessment of college students, and might not. On the contrary, the research program being advocated is exactly the sort of thing one does prior to the development of educational measures and all of the efforts are pursued precisely because they contribute knowledge and materials directly to the assessment effort. Thus, ED may

---

[39] For standard setting procedures, see: W. Angoff, "Scales, norms, and equivalent scores," In *Educational Measurement, ed. R. L. Thorndike* (Washington, D.C.: American Council on Education, 2d ed., 1971): 508-600.

I. Nedelsky, "Absolute grading standards for objective tests." *Educational and Psychological Measurement, 14 (1954):* 3-19; and

R.L. Ebel, "Obtaining and reporting evidence on content validity." *Educational and Psychological Measurement, 16(1956):* 294-304.

For criticism of such procedures, see: G.V. Glass, "Standards and criteria" and others in *Journal of Educational Measurement, 15(1978):* 237-261; and

R.M. Jaeger, "Certification of student competence," In *Educational Measurement*, ed. R. L. Linn (New York: Macmillan, 3d ed., 1989): 485-514.

[40] Ibid., 7.

> want to consider the proposed research agenda itself as the first phase of a larger effort at national assessment. The main purpose behind emphasis on this first phase is simply that, from the standpoint of measurement practice, there are too many uncertainties in the proposed activities to recommend that a particular approach will lead to a valuable assessment, worthy of the public funds that will be needed to support the system in the future.[41]

Dunbar urges national assessment of college student learning developers to actually view his research program as itself the beginning of a national assessment. He believes "the need for a stronger research foundation for educational reform efforts for the next decade and beyond is well established,"[42] and goes all the way back to the 1960s and the birth of NAEP:

> The great educational tasks we now face require many more resources than have thus far been available, resources which should be wisely used to produce the maximum effect in extending educational opportunity and raising the level of education. To make these decisions, dependable information about the progress of education is essential . . . . Yet we do not have the necessary comprehensive and dependable data; instead, personal views, distorted reports, and journalistic impressions are the sources of public opinion. This situation will be corrected only by a careful, consistent effort to obtain data to provide sound evidence about the progress of American Education.[43]

---

[41] Ibid., 23.

[42] Ibid., 26. See, for example M. Kirst, T. James, and L. Shulman, "Forging a national agenda in educational research," *Education Week, 11* (1)(1991): 44.

[43] R.W. Tyler, "The development of instruments for assessing educational progress," in *Proceedings of the 1965 Invitational Conference on Testing Problems* (Princeton, NJ: Educational Testing Service, 1966): 95-105.

Dunbar describes how NAEP both is and is not a comparable assessment, citing as commentators both:

R.E. Stake, R. E. National assessment. *Proceedings of the 1970 Invitational Conference on Testing Problems: The Promise and Perils of Educational Information Systems*). Princeton, NJ: Educational Testing Service, 1971): 53-66; and

R.L. Linn, *Historical origins and issues in the National Assessment of Educational Progress.* Paper presented at the Institute for Practice and Research in Education forum on Assessment at the National Level (Pittsburgh, PA, October, 1990).

"Devise more pragmatic and immediate systems"

Peter Ewell and Dennis Jones, like Dunbar, appreciate the magnitude of undertaking a full-blown national assessment of college student learning, and contend that direct assessments of critical thinking, problem solving and effective communications "are technically complex and will take many years to develop."[44] Why not collect and examine what they describe as "indirect indicators" in the meantime? They do not contend that such measures will "substitute for the kind of purpose-built, performance-based"[45] national assessment of college student learning being contemplated, but they do point out that the technical properties of assessments (like the major, instrument-based national assessment of college student learning under consideration) "are complex and in many cases unknown; standard validity and reliability measures are hard to apply and the results are often subject to unknown and uncontrollable biases."[46]

By their choice of the term "indirect indicators," Ewell and Jones do not mean to concede that such measures might not be as useful as other types of indicator, given the still murky line of logic in Goal 5:

> It must be emphasized that all indicators of educational attainment are in some sense indirect. Purpose-built, performance-based assessments of particular areas of knowledge and skill, such as that proposed by the National Goals Panel for tracking progress in collegiate attainment, are no exception and like the results of any test, should not be confused with the actual entity that they purport to represent.[47]

In making this distinction, they clash head-on with Lenth who forcefully argued the value of outcomes measures. They describe recent trends in American education to pay "greater attention to the outcomes of higher education over the last five years, both at the state level and in accreditation."[48] But they do not concede that outcomes indicators are inherently better at guiding actual improvement. Thus, with the *instruction and attainment* of underlying Objective 5 abilities in mind, rather than a clear and testable direct *demonstration* of them, Ewell and Jones propose a series of indicators—most of which are already in existence—to reveal the extent to which critical thinking is manifest in the instructional process of American education.

---

[44] Peter T. Ewell and Dennis P. Jones, "Actions Matter: The Case for Indirect Measures in Assessing Higher Education's Progress on the National Education Goals" (paper commissioned for the NACSL study design workshop by the U.S. Department of Education, National Center for Education Statistics, 1991): i.

[45] Ibid., 1.

[46] Ibid., 3. See, for example: C.R. Pace, *Measuring the Outcomes of College: Fifty Years of Findings and Recommendations for the Future.* (San Francisco: Jossey-Bass, 1979);

E.T. Pascarella, and P.T. Terenzini, P.T. *How College Affects Students: Findings and Insights from Twenty Years of Research.* (San Francisco: Jossey-Bass, 1991).

[47] Ibid., 6.

Indirect indicators of "undergraduate attainment are of quite different types," explain Ewell and Jones. But they all "embrace what colleges and universities *require* of their students, what typically *happens* in a collegiate classroom or course of study, and what colleges *do* as a part of their education."[49] For example:

- ○ *Institutional Requirements* such as specific proficiencies, types of experiences and "capstone" or other integrative experiences required for graduation.

- ○ *Instructional "Good Practice"* involving principles such as active learning, frequent feedback on performance, and the frequency of contact with faculty. Examples: class size in lower-division courses; instructional experiences that entail writing, speaking, explicit problem solving, independent work and essay- or problem-type final examinations; out of class experiences, such as contact and participation with faculty, group and independent study, off-campus work, tutoring others; institutional policies or investments, such as class size, what faculty teach which undergraduate courses, and how faculty are rewarded for doing so.

- ○ *Student Behaviors and Self-reported Gains*, such as how students spend their time in certain selected areas, how they perceive their own gains in the target skills, and how they report their own reactions to college-level work.[50]

Ewell and Jones emphasize that their plan only provides examples of the indirect indicator approach, which are best used in combination, and require further study as to their feasibility.[51] ***Elinor M. Greenberg's*** response to the "Indirect Indicator" approach was dynamic, as she was on first reading "tempted to reject" the plan and its rationale. But she eventually changed her mind, and explained why:

---

[48] Ibid., 7.

[49] Ibid., 10.

[50] Ibid., 10-13.

> The paper is both thoughtfully prepared and pragmatic in approach. It makes a good case for using what is already available in order to influence policy, at least in the short run, while keeping an eye on *improvement* as the driver for assessment efforts, in the long run. If, as the authors propose at the outset a *few* indirect measures are used *concurrently* with the development of other efforts, the institutions are more likely to "buy in" earlier and may see the overall national assessment task as more friendly and doable. The immediate availability of data sources is the main advantage of this approach, while the inferential nature of those sources is the main disadvantage, relative to assessing higher order thinking and communication skills.[52]

Greenberg decided that the data to be collected in the Ewell/Jones plan could provide a necessary benchmark for her own view of what is needed, "a multiple option assessment plan, which could be implemented *over time* (1992-2000, eight years that cover two complete 4-year baccalaureate cycles)," which she calls the "Institution-Based Assessment Option." This plan would embody the "continuous improvement" approach, and include performance-based assessments consisting of Development-based [see Loacker paper], Industry-based [see

---

[51] Ewell and Jones suggest development along several lines, if an initial exploration of their scheme seems promising:

o A systematic review of currently available national data collection instruments and methodologies. This to determine: if their information is adequate to the proposed, to-be-defined reporting purpose; if the sampling basis allows valid generalizations to national populations; and/or if necessary and appropriate modifications for these purposes can be made.

o A major background paper on the validity of indirect indicators, which by a systematic review of the literature on outcomes might establish or identify linkages between the process and "proxy" indicators and the desired outcomes; i.e., students' abilities to think critically, communicate effectively, and solve problems.

o A feasibility study designed to determine the costs and logistics involved conducted on a typical sample of institutions and respondents.

The authors see this as a two-year developmental process before appropriate national indicators could be put in place, but nonetheless define it as a short-term alternative. Its value thus would be to keep the national policymakers focused on postsecondary education, as well as to take advantage of a generally-held belief, that "such indicators are generally of greater value in inducing institutional change than are outcomes indicators used alone. On balance," they conclude, "it seems a path worth exploring further." [p. 25]

[52] Greenberg on Ewell and Jones.

Cappelli paper], and State-based [see Morante, Lenth papers] efforts "which have already begun and are likely to proceed, in any case." In this recognition she echoes Lenth's sensitivity to encroaching on current turf and to competing with extant assessments. She would build on education/industry assessment partnerships now being called for and developed, nation-wide. Then, cooperative arrangements between the Department of Education and the Department of Labor could provide the national structural framework, using data now available to each Department. This kind of inter-agency collaboration would begin to build a partnership type of structure at the federal level that could be very useful well into the 21st century and would be a model for partnerships in the field.

> My guess is that within 2-5 years, these initially discreet efforts would converge, giving us a more coherent national view of performance while, at the same time, building a national consensus about the competencies needed for the new workforce in a global marketplace and how to achieve them. "Good practices" would need to be more widely publicized in order to be understood, accepted and, hopefully, utilized. At many levels, we already know what is wrong, but we've been unwilling to do the inconvenient work it will take to fix things.
>
> This indirect, institution-based approach, then, is a conservative and pragmatic one, not designed to do the whole job of national assessment and improvement, but intended to start the process without undue delay and controversy.[53]

The four options listed above would provide a choice, for all U.S. institutions, among systems tilted towards one or another particular structural bias: industry, development, state, or institution, "thereby creating four approaches that could be compared as to effectiveness and efficiency."[54] Greenberg styles this "total effort [as] the Coordinated Multi-Option National Assessment and Partnership System:

> If phased in and continued from 1992 to 2000, this eight year period would capture two traditional four-year baccalaureate cycles and bring us to "The Class of 2000" with a substantial national database, as well as with a de-centralized and diverse national assessment system. This pluralistic approach could satisfy the various constituencies and stakeholders who are now and will continue to be major players in

---

[53] Ibid.

[54] Greenberg on Cappelli.

this process. It would also be "grass-roots," "team-based," "customer-driven" and focused on "continuous improvement," matching the Total Quality Management (TQM) approach, currently so credible with industry and government leaders who are working to improve American productivity.[55]

Greenberg calls for a "perspective transformation" as the only way to change current institutional practices:

> This is not easy to produce. Urgency, not patience, is the order of the day. The credibility of higher education to help increase workforce productivity in the global marketplace, as well as the quality of our citizen-based democratic process, is, indeed, at stake. Therefore, I urge us to adopt a more dramatic, multiple option, collaborative, far-reaching strategy of which "indirect" measures, or an institution-based approach, is but one option among many, and one that can begin immediately.
>
> We might think about this entire national assessment effort as a decade-long movie. This "indirect" portion is a series of snapshots that only tells us part of the story, but it holds too still to capture the true dynamic of lifelong learning. At the same time we are looking at the frame that shows us what is happening now, we must run the reel forward, if we are to catch sight of a better future.[56]

Thus her call for a "Collaborative Multi-Option National Assessment and Partnership System." To begin to develop this system, she suggests, "the Department of Education, and especially NCES, would be required to play a consultative and coordinating role in the development of a large "Network Organization" of educational providers. This would require some shift in role and strong leadership in communicating that shift, along with its value and timeliness, to institutions, other federal departments, Congress, the President, the Governors, the states, and the public."[57]

---

[55] Ibid.

[56] Ibid.

[57] Ibid.

Given the diversity of responses to the NCES charge, it might be asked, Was there one idea or response common to all of the participants? Looking beneath the often forceful caveats and the serious apprehensions and misgivings many expressed about national assessment of college student learning, perhaps a unanimous vote could be achieved on hope. It wasn't blind, however, and the strong convictions many held about ***not*** *proceeding straightaway with the development of a single instrument were often based on the belief that it simply wouldn't work. That is, would not achieve its avowed purpose, and—having scored a political and institutional failure—would set back long-term prospects for what all participants did agree on: that critical thinking should be fostered in American education.*

Daniel Resnick and ***Natalie Peterson*** exemplified this attitude, focusing not so much on how to quantify the increase in critical thinking among American college graduates as on how to enhance and encourage it. Their own criteria suggest that measures should "not only monitor progress toward the goals, but help to make sure that the goals are achieved."[58] They emphasize that "neutral scientific measures, though they should be valid and reliable,"[59] are not likely to effect these policy goals. Looking behind the language of Goal 5, they believe the essential problem could be stated differently, they say, citing *SCANS* as well as another relevant commission study:

> Since the standard of living of the nation depends increasingly on bringing to the United States and keeping in this country a large number of highly skilled jobs, the national policy agenda is open for proposals to raise the level of education of school graduates and to upgrade capabilities of already employed workers."[60]

Thus a more direct way of realizing Goal 5, they suggest, is to realize that public policy is somehow going to affect two vital communities of interest: American public education (at all levels) and the American business community. This, continue the authors, is because

---

[58] Daniel P. Resnick and Natalie L. Peterson, "Evaluating Progress Toward Goal Five: A Report to the National Center for Education Statistics," (paper commissioned for the NACSL study design workshop by the U.S. Department of Education, National Center for Education Statistics, 1991): 1.

[59] Ibid., 2.

[60] Ibid., 2. National Center on Education and the Economy, Commission on the Skills of the American Workforce, *America's Choice: high skills or low wages!* (Rochester, NY: National Center on Education and the Economy, 1990).

> the American political agenda is to create a workforce with higher order skills capable of: solving problems, recognizing the broader context of workplace activity, learning how to keep on learning, using technology effectively and devising adaptive responses to new challenges. These are challenges that have been set before both the population that enters the workplace directly after high school, and the segment that enters after completing one or more years of college.[61]

Their reframing of the thrust of Goal 5 in more pragmatic terms leads them to

> recommend that the Department of Education employ a variety of indicators. Multiple measures will be needed to convey the changes in our institutional functioning and cultural expectations that are to necessary to measure the progress of educational reform. The indicators and reports that we propose the National Center for Education Statistics provide were chosen to enlarge the public's understanding about the many factors that affect change in the educational achievements of our population.[62]

Most of their recommendations are considered in subsequent chapters. The six, in brief, and the questions they are designed to answer:

1. The best way to assess learning gains in the secondary schools?
   - Monitor the availability of AP courses at the individual high school level, how many students take such courses, and how they perform.
2. The importance of demanding curricula in major fields of college learning?
   - Conduct a transcript survey of a selected 30-50 institutions modeled after that conducted by Zemsky and his colleagues at the University of Pennsylvania.

---

[61] Ibid., 2.

[62] Ibid., 21.

3. The role that can be played by disciplinary associations in changing patterns of college study?

    - Issue an inventory, broken down by field of major study, of guidance promulgated by the various professional institutions.

4. The ways in which colleges and universities are establishing learning outcome measures for their students?

    - Encourage demanding standards in major programs, focusing on capstone senior experiences that might include comprehensive written and oral exams, theses, projects, portfolios, self-reports and a variety of demonstrations.

5. The gulf between workplace needs and the perceived capabilities of recent school graduates?

    - Model a survey after the Harris 1990 version to gauge the extent to which recent college graduates are able to meet the needs of today's workplace.

6. The need for continued monitoring of literacy-related skills in the adult population?

    - Expand the NAEP Young Adult Literacy Assessment so that recent college graduates can be disaggregated for analysis.

All of the foregoing primary authors addressed the national assessment of college student learning challenge on the assumption that the intellectual exercise they were participating in was the overture to some sort of new—but very real and increasingly likely—educational event. Each had come from a particular experience, and used that background to try to emphasize aspects of development they thought crucial. To recapitulate:

- ***Paul and Nosich*** want to be certain that the conception of critical thinking which finds its way into the national assessment of college student learning remains as true as possible to the state of the art in that field.

- ***Rock*** wants to clarify lessons learned from previous testing experience in large-scale testing at both the postsecondary and lower levels, and hopes that some possible but predictable mistakes might be avoided.

- ***Lenth*** understands that the people on whom the success of the project depend are already known: they are the educators and policy people who run the educational infrastructure in the 50 states. It is crucial to understand what they may need, what they will be willing to adopt and support, and what would be both redundant and/or threatening to them and their current institutional world.

- ***Morante*** knows how a major, state-wide test gets developed, implemented, conducted, used, and how it may fare politically. The New Jersey experience, though obviously not a pure fit to the nation at large, is nonetheless the best overall functional model we have.

- ***Cappelli*** appreciates that most students ultimately being assessed are going to have to find their way into the workplace, and wants both camps to begin looking into this eventuality sooner in a student's career, and hopes a better transition might be accomplished through a national assessment of college student learning.

- ***Venezky*** knows literacy assessment, and wants to be certain that what has been learned from that domain not only be respected as it is translated into the national assessment of college student learning, but that it be considered for lessons about the test itself.

For all of these authors, an instrument was more or less accepted as the premise, and all dealt from their own particular perspectives with the implications and complications of how such an instrument might be designed. Even Dunbar's forceful warnings to slow down and be rigorous about validity and consequentiality shared as a premise the eventual creation of an instrument, however further innovative analysis and insights may reveal that its structure and application needs to be. The two author/teams already mentioned produced papers suggesting substantive possible plans for proceeding *now*, which did not, however, include the construction and implementation of a new, monolithic instrument. Ewell/Jones and Resnick/Peterson each suggested creative ways to address Goal 5, with different measurement strategies that might precede (or even obviate) the creation of a dedicated major instrument.

The six remaining primary authors, however, could not and would not be constrained by that assumption. Nummedal, Ratcliff, Banta and White each—like all of the authors —possesses a unique and valuable experience that could inform the development of a national assessment of college student learning. In the end, none of them would buy into the concept of a major national one-time test, each for reasons that they explained and rationalized in detail—reasons they believe suggest a re-evaluation of where we are and what we are doing.

"Improving instruction should be the major premise"

Trudy Banta from the University of Tennessee goes perhaps the furthest of those who would respond to Goal 5 with a dramatic vision for change. She lays the foundation for her proposal by discrediting the task as it has been structured by most participants in the dialogue. Perceiving five basic assumptions implicit in Objective 5, she proceeds to unmask each of them, laying bare how at odds their implications are with her take on the realities of American educational assessment. Her first assumption questions whether the Objective 5 abilities can be defined and agreed upon, and is taken up in some detail in Chapter 2. The other four—all of which she believes are unfounded in practical or historical reality—lay the groundwork for her own reform proposal.

2. The defined abilities will be taught, by all faculty charged with the responsibility for teaching them, in ways that engage students and promote learning of these abilities.

3. Reliable and valid measures of student achievement of the defined abilities can be identified or created.

4. The measures of student attainment can be administered to all college graduates (or samples of that population) in settings that engage students and encourage their best efforts.

5. The **results** of assessment of developed student abilities will be **used** to **improve** the materials and methods of **instruction** in ways that increase student engagement and **promote learning gains**.

This fifth premise is her centerpiece:

> Given my deep personal commitment to the improvement of practice, the emphasis throughout this paper is on the **process**—what evidence do we have that the current process of assessing postsecondary student outcomes is capable of producing improvements in those outcomes; and if that process is inadequate, how might positive change be effected?[63]

While a number of participants agreed about the importance of instructional reform, Banta strikes hard at current conceptions of the national assessment of college student learning as inimical to a process that could actually have a constructive impact. "Some of the most knowledgeable measurement specialists say that it is not currently possible to develop an assessment program that meets the twin goals of monitoring status for accountability purposes and providing direction for instructional improvement because optimizing validity for one purpose diminishes it for the other."[64]

Banta doesn't suggest that monitoring by assessment has no place—

> Monitoring progress, or assessing status, is a component of any effective process. But if there is anything that Edward Deming,[65] Japanese industrialists, and winners of the Malcolm Baldrige Award in this country have taught us in recent years, it is that inspection alone will not produce improvement.[66]

---

[63] Trudy W. Banta, "Toward a Plan for Using National Assessment To Ensure Continuous Improvement of Higher Education," (paper commissioned for the NACSL study design workshop by the U.S. Department of Education, National Center for Education Statistics, 1991): 3.

[64] Ibid., 14. She cites P.A. Moss, and S.M. Koziol, "Investigating the validity of a locally developed critical thinking test," *Educational Measurement: Issues and Practice*, 10(3)(1991): 17-22, and quotes G.R. Hanson's belief that "assessing when and how students change, and linking such change to specific educational interventions, is a complex and difficult task that requires new strategies for conceptualizing issues, building new and different assessment instruments, and designing research with different purposes and outcomes than those found in many traditional methods of inquiry." G.R. Hanson, "Critical issues in the assessment of value added in education," in ed. T. W. Banta, *Implementing outcomes assessment: Promise and perils. New Directions for Institutional Research*, vol. 59, (San Francisco: Jossey-Bass, 1988): 54.

[65] E.W. Deming, *Out of the crisis*. (Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study, 1986).

[66] Banta, op. cit., 14.

but rather that it must serve the more central goal of improvement. Banta believes that "we must *secure* that investment by making the monitoring activity part of a larger system that ensures the *use* of assessment findings to *improve* education."[67] To establish the permanent system she believes that "we must:

- specify clear goals and objectives for the skills we seek in college graduates,
- provide the staff development and instructional resources necessary to prepare faculty to teach these skills using methods that genuinely help students learn them,
- develop precise measures of the specified skills and administer these to students in ways that encourage their best efforts, and
- use the results of assessment to modify the components of this system that are shown to be in need of improvement.[68]

"This will require a national effort of epic proportions," she concludes. "It will be enormously costly. But if we are determined to attack this problem, and to do so in ways that have a chance of being effective, we must begin systematically, drawing upon everything that recent experience with assessment at elementary, secondary, and postsecondary levels has taught us."[69]

How to actually implement such a revolutionary reform? Banta outlines an approach that she believes would build upon the current national assessment of college student learning energy, but one which would clearly involve much more groundwork, developmental research, and reorientation than is being contemplated by most planners. Begin by empaneling in each of the states and at the federal level what she calls "Objective 5 Panels," whose first task would be to consult "previous efforts to define a domain of knowledge,"[70] and who would select the skills and abilities according to an overarching standard: "what a competent adult should know and be able to do in each of these areas."[71] Such a draft proposal then serves to trigger an elaborate and extensive dialogue with interested public, and "especially from faculty at public and private colleges,"[72] possibly structured with the Delphi process.

---

[67] Ibid., 15.

[68] Ibid., 15.

[69] Ibid., 16.

[70] Ibid., 17. See for example: C. Adelman, "Introduction: Indicators and their discontents." In C. Adelman (Ed.), *Signs and traces: Model indicators of college student learning in the disciplines* (Washington, D.C.: U.S. Government Printing Office, 1989): 1-10.

To Banta, faculty participation—as it is for Ratcliff, Lenth, and others—is paramount. Since improving instruction would be the focus of her program, she believes it crucial that "every college and university faculty should decide upon its own *program of in- and out-of-class experiences* that will promote student development of the specified abilities. Selected faculty and staff should be charged specifically with the responsibility for providing these experiences in courses and out-of-class activities."[73] Banta sees this as the province of teams of staff development specialists who would be responsible state-wide to acquaint college faculty with the teaching strategies and materials that had been developed at the federal level. But such a process would be no rubber stamp:

> Continuous student and faculty review and evaluation of teaching strategies and materials must be built into this process. And as experience proves certain approaches to be more valuable than others, this information should be used to modify the curriculum used by the state staff development specialists.[74]

Trudy W. Banta, *The competent college student: An essay on the objectives and quality of higher education* (Nashville, TN: Tennessee Higher Education Commission, 1977).

P.A. Facione, Executive summary of "Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction." (Millbrae, CA: California Academic Press, 1990).

Alverno College Faculty. *Assessment at Alverno College.* (Milwaukee, WI: Alverno Productions, 1979).

D.W. Farmer, *Enhancing student learning: Emphasizing essential competencies in academic programs* (Wilkes-Barre, PA: King's College, 1988).

G.W. Peterson, *A meta-evaluation of a generic skills approach to the evaluation of academic programs.* (ERIC Document Reproduction Service No. ED 219-398, 1982).

[71] Ibid., 17.

[72] Ibid., 17.

[73] Ibid., 19.

[74] Ibid., 19.

Once the basic goals and structure of the program had been certified by this extensive dialogue with the institutions, then would come the time to consider assessment strategies, and again Banta proposes to establish a different "federal panel of outstanding measurement specialists and college and university faculty."[75] The first crucial task of this panel would be to gather information about the underlying activities that describe individual campuses, programs, departments, and students, and Banta points to some likely sources for such data:[76]

1) Is student growth a clearly articulated and implemented institutional goal?

2) Is each student and faculty member aware of the federal expectations with respect to student development of Objective 5 abilities?

3) How much time has each faculty member spent in staff development activities related specifically to promoting students' learning of Objective 5 abilities?

4) How much time does each faculty member spend preparing to teach, and teaching, material related to Objective 5?

5) How much time does each student spend studying material related to Objective 5?

6) How much out-of-class time does each student spend in conversation and/or activities related to the Objective 5 abilities?

7) Do students and faculty perceive that they have access to the facilities, equipment, experiences and materials they need to promote development of Objective 5 abilities?

8) Is student progress toward development of Objective 5 abilities sufficiently evaluated, and is the student briefed concerning that progress?

9) Are students sufficiently motivated to develop the Objective 5 abilities and to do their best work when their progress is evaluated?

---

[75] Ibid., 20.

[76] Ibid., 20.

Banta forecasts difficulty in discerning such assessment data with the major instrument approach. As a practical alternative, she believes that a more performance-based system might evolve that emphasized a senior project, echoing the capstone approach suggested by Ewell and Jones. Nancy Beck remarks that Banta says nothing about "equating the senior activities. And the problems of doing so, given the likely diversity of the activities, make it unlikely to occur. Without some form of equating and/or comparability of activities across students/institutions, it is hard to see how any trend data could be established."[77] Banta concedes that while enormously more complex and costly to evaluate than an objective test, such a structure would nonetheless provide the sort of baseline that could facilitate comparisons between institutions and possibly states.

Frederiksen believes such methods rely on a form of embedding assessment activities in coursework which he believes "a sound idea:

> The teacher begins by posing a problem to students, who are encouraged to form small groups to work together. Help can be provided as needed in the form of hints, reference books, models, computers, video, teacher aides and teachers, etc., as necessary. As problems are solved, more difficult complex problems can be presented. Understanding of the domain grows as success in the earlier tasks provide a background for further learning and mastery. As the term of teaching continues, records of the performance of each student can be preserved and used as a basis for assessment.[78]

Banta is not insensitive to the sweeping scope of her proposal, nor to its financial and institutional implications. Faculty in America would doubtless feel they were entering a brave new world of educational adventure (the peril of which they would most likely apprehend in proportion to their satisfaction with the current state of affairs), and would encounter two major dilemmas: some of their current activity would have to be supplanted by staff development, planning, data-gathering and analysis activities, and once the core competencies were established and promulgated, academic freedom for individual faculty could be affected.[79] Notwithstanding these caveats, Banta rings forcefully her appeal to "do it right . . .."

---

[77] Beck on Banta.

[78] Frederiksen on Banta. He believes "such procedures have been found to produce results that are far superior to the blackboard-and-eraser lectures. (See any copy of a new journal named *Interactive Learning Environment*, Ablex Publishing Corporation, 355 Chestnut Street, Norwood, NJ 07648).

> If the decision is made to invest the national resources necessary to carry out the comprehensive approach to assessment-and-improvement proposed in this paper, the Secretary of Education would have an opportunity to establish the boldest and potentially most promising research and development project ever undertaken in higher education. This work could establish the basis for making continuous improvement a part of everything that is done in the name of postsecondary education. This development could help colleges and universities reclaim some of the responsibilities for providing higher education that they are losing to private industry and federal agencies. Finally, this approach could ensure that the higher education system in the United States will be sufficiently responsive to changing global needs to maintain its current reputation as the best in the world.[80]

Nancy Beck's review puts this clarion call into perspective:

> While potentially attractive from an educational point of view, the proposed system would be impossible to fund or carry out. In spite of this generally negative review, [Banta's] paper could be useful to policymakers. She very effectively shows the implications of taking a single goal/objective at face value and carrying it to its limit. What seems more likely is that policymakers want something that will help turn the ship while recognizing that they cannot reform higher education with one objective of an overall set of broad educational goals.[81]

---

[79] Banta, op. cit. 29. See, for example, M.A. Miller, Assessment in hard times. *Assessment Update*, 3(6)(1991): 1,3,5.

[80] Ibid., 29.

[81] Beck on Banta.

"Real life relevance should be the engine"

Susan Nummedal carefully distinguished among various and sundry definitions of skills/abilities in order to lay the groundwork for a fairly strong recommendation,[82] not simply of which skills to target, but how to frame the entire national assessment of college student learning. Nummedal notes that in the *Authors' Briefing Paper* a kind of "ambiguity" exists which is really a function of the NEGP report and, indeed the entire dialogue to date.

> Does this mean the assessment should address both the higher order thinking skills that are or should be acquired through the college experience **and** those associated with real world success **or** only those higher order skills [which are] acquired though the college experience that are associated with success in the workforce and with effective citizenship?
>
> I want to argue that we should focus our attention on the latter, namely those skills acquired through learning experiences in college that are relevant to successful functioning in real-life situations.[83]

Thus while her arguments are mentioned in both subsequent chapters under choosing skills and abilities and establishing standards for such skills, the basic argument she makes goes deeper. Transfer [successfully into the workplace] is not merely one aspect or measure of a chosen skill, but is, she believes, a major premise to identify skills that should drive the entire national assessment of college student learning process. The background she provides on how to discern such skills is described in Chapter 2, but the distinction she makes has far-reaching implications that would essentially derail the push for an instrument.

> This discussion of everyday thinking points to two important implications for assessment. First, implicit in the characterization of everyday thinking as being flexible and efficient is the notion that people have available multiple strategies for solving problems *and* that the first one(s) selected may often represent something *less than their 'highest' level of thinking* about solutions to the problems at hand. If we give a person just 'one shot' at solving a problem or addressing an issue (even if this

---

[82] Susan G. Nummedal, "Designing a Process To Assess Higher Order Thinking Skills in College Graduates: Issues of Concern," (paper commissioned for the NACSL study design workshop by the U.S. Department of Education, National Center for Education Statistics, 1991): 3.

[83] Ibid., 4.

> 'shot' is under the guise of a *National Test*) we need to be very careful about what we conclude that person knows or does not know. To more closely approximate everyday thinking and problem solving, a performance based assessment needs to be a dynamic process.[84]

Moreover, the special qualities she believes are required for critical thinking that will prove to be effective in the workforce would require extremely creative and innovative approaches to measurement, a conclusion which anticipates themes to be considered in the discussion of the system at Alverno College.

> This notion of feedback is directly tied to [the fact] that everyday thinking and problem solving most often occurs in interaction with others. Thus, one type of assessment that needs serious consideration is assessment within a group context. Asking students to solve problems in conjunction with others would more nearly model real world problem solving. Since this assessment is not designed to provide information about *individuals*, there is no reason to reject outright a group problem-solving exercise.[85]

Nummedal is presently Chair of the Assessment Group of the California State University Critical Thinking Council. As such, she might have been expected to echo and reinforce the critical thinking credo and the call for a major role for critical thinking experts presented by Paul and Nosich, Facione, and others. But her survey of critical thinking instruction and her first-hand knowledge of the California experience does not persuade her that the critical thinking establishment in and around American colleges—in its current state—would be a sufficient foundation for a truly successful national assessment of college student learning, one that would be demonstrably and intellectually sound.[86] And, even more importantly, critical thinking in its present state is not sufficiently coherent to produce a reliable recommendation about how to accomplish what Nummedal considers a fundamental premise of the drive underlying a national assessment of college student learning: improving instruction.

---

[84] Ibid., 8.

[85] Ibid., 8.

[86] Ibid., 10-11. See also Chapter 3 for a discussion of validity.

"To improve instruction, first you must locate it"

When we talk about an assessment designed to inform practice, one question we must ask ourselves is, "*Where is the instruction* in critical thinking, problem solving and communication that we want to inform *taking place*?"[87]

To begin to answer this question, Nummedal points out that it is difficult "to talk about 'informing instruction' as if instruction were a unitary thing:

> Critical thinking could be cultivated in courses designed to teach such skills exclusively and explicitly; in courses (such as general education) where critical thinking skills are but one of several goals for the course; in courses where critical thinking is not explicitly taught, but where its practice is essential to success and is an inherent part of the way the course is taught; or within the major in the form of phenomena like courses in research methods.[88]

Her first thought about how to realize this new direction is that a "typology of institutions" might be developed to locate within a given university the above-listed potential sources of critical thinking instruction, and to quantify an individual student's curricular exposure to them. As "evidence that these goals are being addressed,

> . . . Content analyses of college catalogues, general education guidelines, and other official documents might prove useful. These, however, are relatively distal measures of these goals. More proximal ones might include examining course syllabi, assignments, and examinations used in specific courses. Or one could analyze student products for evidence of the implementation of these goals. One could further refine this approach by assessing the educational context for individual students based upon course patterns within types of institutions.[89]

---

[87] Ibid., 13.

[88] Ibid., 13. For clarifying a fifth and more global possibility, the general university experience, she credits W.G. Perry, Jr., *Forms of intellectual and ethical development in the college years: A scheme.* (New York: Holt, Rhinehart, and Winston, 1970).

[89] Ibid., 14.

She acknowledges that such an analysis may not achieve the avowed goal of immediately measuring current critical thinking abilities, but her entire argument indicates that such a measurement is neither feasible nor advisable. Hambleton concedes her point: "I agree very much with her ideas for improving aspects of instruction and assessment at the college level. I doubt, however, that a national assessment system could evolve from her work. Of course, now the dilemma is clear: Why teach critical thinking one way in the classroom and then have a national assessment system which is only a crude proxy for what is really important?"[90] There is no short cut to improving Americans' skills in critical thinking, Nummedal contends, and the assessment movement could—if it were re-framed—provide a legitimate foundation. This assessment, she argues, should be

> about informing practice and practitioners of all stripes—students, educators, employers, assessors, and policy makers. If we see the primary purpose of this assessment as *informing practice*, then we will it as the start of a dialogue *with*, not a report *to*. We will see the assessment as a dynamic process which engages all the players.[91]

As an example, she envisions college instructors at the nexus where information and communication meet, in a position to use "their expertise. . .to help shape the assessment tasks and guide the process of setting performance standards. So too for other professionals."[92] Not until this dialogue has begun to clarify and refine the difficulties can anything like a definitive assessment process be developed.

> We will see it as an ongoing process. There are far too many questions to which we have no answers. We are far from a full understanding of the nature of higher order thinking skills. We are far from understanding how the educational process might bring about the kinds of changes in higher order thinking we are trying to assess. We might do well to think of this assessment as an opportunity to discover what some of

---

[90] Hambleton on Nummedal.

[91] Nummedal, op. cit., 14.

[92] Ibid., 15.

> the interesting questions about this process really are rather than as a final report on 'what is.'[93]

Nummedal's suggestion, it turned out, was to anticipate an even more substantive alternative plan for the national assessment of college student learning proposed by ***James Ratcliff***.[94] He echoed some of Lenth's concerns with his belief that

> For a national program assessing college graduates to succeed, it must be institutionalized at the majority of colleges and universities in the nation. The information generated from such an assessment must play a clear and key role in formulating state and institutional higher education policy, in college matriculation standards and degree-granting decisions. In short, it must affect the teaching of faculty and the learning behavior of students. For colleges and universities to give such meaning to an assessment program, it must be credible, reliable and useful.[95]

Ratcliff [like Dunbar, Nummedal and Banta in particular], if he were going to overturn the basic premises on which the national assessment of college student learning exercise is perched, felt obliged to document a strong intellectual rationale and alternative.[96] He places the concern for quality in higher education in the context of recent American experience, and sees therein a common thread. Like the national reports[97]

> that recommended higher education standards be raised, the curricular prescriptions for students be changed, and the content and structure of degree requirements be fortified, the objective of the National Educational Goals Panel (NEGP) relative to

---

[93] Ibid., 14.

[94] James L. Ratcliff, "What Type of National Assessment Fits American Higher Education?" (paper commissioned for the NACSL study design workshop by the U.S. Department of Education, National Center for Education Statistics, 1991).

[95] Ibid, i.

[96] In brief, first his explanation about why we can't get there from here [to be followed by his proposal for a new direction, which fleshes out Nummedal's "typology" considerably].

[97] Ibid., 1. "The higher education reform proposals of the 1980's made implicit assumptions about what constituted effective undergraduate education. The three most frequently cited of these, *Involvement in Learning* (NIE, 1984), *To Reclaim a Legacy* (Bennett, 1984), and the American Association of Colleges' *Integrity in the College Curriculum* (1985), cited a decline in the quality of liberal or general education and called for reforms to strengthen undergraduate programs."2

> college graduates assumes that there is a common body of knowledge, skills and abilities to be imparted through an undergraduate education, and that there are readily available means to assess student progress in acquiring that common knowledge. Yet, our current ability to make such assessments is limited by the relative lack of common curriculum and relatively limited range of means and measures of that curriculum.[98]

Will the national assessment of college student learning (Ratcliff prefers to characterize it as "a NAEP-like examination") improve instruction? He says there is no "widespread institutional and faculty sense that such testing would" do so. Moreover, the call for a national assessment "presumes that commonly accepted means and methods are available to assess student growth in critical thinking, communications and problem-solving abilities. The call presumes not only a consensus of criteria on what constitutes learned abilities in these three areas but also assumes that, given the creation of an appropriate yardstick, higher education will have the means and the resources to improve student performance."[99] Such presumptions, believes Ratcliff, are unwarranted and drastically premature; as did Banta, he demonstrates why.

Especially at risk are some elements of postsecondary education that did not complicate NAEP at lower levels. [Though the question of student motivation addressed by most authors is being considered in a subsequent NCES exercise.] Ratcliff believes the rigid structure that a monolithic test would impose clashes with both the voluntary nature of postsecondary education and with the advantage to American society in attracting students of diverse backgrounds and capabilities. He cites the experiences in Florida and Texas as a warning. He takes it as given that students must be encouraged, and does not believe that assessment *per se* is discouraging. But he *does* favor tests which are constructed for their diagnostic value in helping college students to "choose coursework which [is] challenging and appropriate to their ability level."[100]

The problem inherent in a national test, many commentators agreed, is the inevitable political use(s) to which it will be put. Comparison between institutions promotes a kind of competition in the heat of which the interests of individual students can be singed. This is unfortunate, notes Ratcliff, because "the major differences in student abilities [are found] among students on the same campus rather than between students from different campuses. The real challenge, he believes, is to use curricular and assessment information to help students select coursework that is appropriate to their interests, abilities and skills."[101]

---

[98] Ibid., 2.

[99] Ibid., 8.

[100] ibid., 12.

This recommendation builds on the advice of Lenth, Morante, and others whose chorus about bringing value to all players makes good political sense. And, predicts Ratcliff, good psychological and common sense as well: "Assessments in [the] service of finding proper matches between student abilities and educational environments are far more likely to produce gains in student learning than are those that seek to discourage and impede the progress of the less able student."[102]

Another major distinction between postsecondary and lower educational systems is the freedom of faculty to participate in implementing policy. "*Involvement in Learning* (1984) recognized the importance of using faculty in the assessment of student learning,"[103] reminds Ratcliff.

> A first step in that process is to establish the credibility of the assessment program in the eyes of college faculty. A second step is to find ways to involve them directly or indirectly in the assessment process. In doing so, the assessment program also insures that faculty understand fully what is being assessed. Faculty are in the best position to draw students' attention to the outcomes their college or university values the most. As the *Involvement of Learning* report indicates, assessment has even greater potential as a tool for clarifying expectations and for increasing student involvement when it is used to measure improvements in performance.[104]

Ratcliff looks more favorably on the California experience [than did Nummedal] as an example of properly inducing faculty to "buy into" the assessment process.

---

[101] Ibid., 12.

[102] Ibid., 12.

[103] Ibid., 13.

[104] Ibid., 13.

> An example of how faculty and institutions can be encouraged to incorporate an assessment program in the teaching, learning process is the California Basic Skills Instruction policies of the California Community College system. The California Precollegiate Basic Skills Instruction policies require colleges to establish ability level prerequisites for degree-applicable, entry-level courses. These policies require institutions to develop standards for the rate at which a student progresses toward a degree and limits students to earning a maximum of 30-semester-units in precollegiate basic skills courses. Each institution must define the scope of their student assessment program and relate them to student course selection. A statewide task force examined the costs of the implementation of these policies to the individual institutions and to the state.[105]

Since students arrive at many public universities with a widely varying background, and then proceed to choose from an extraordinary range of course selections, a one-time assessment improperly interpreted could obscure rather than reveal the underlying phenomena that assessors are ostensibly trying to locate and improve.

Funding struggles, finger-pointing over accountability, "the journalists' pension to administer guilt by association"[106]—all of these political realities prompt Ratcliff to propose a revolutionary system of assessment that would focus, not on institutions, but rather on states, and on individual students. He illustrates by example. Consider

> the university with 5,000 courses in which undergraduates may enroll. Let us suppose that our assessment shows an overall decline in problem-solving ability. To improve the educational program, it is a foolhardy waste of resources to insist that all 5,000 courses increase their focus on problem solving. First, not all students in the group tested took all the courses. Secondly, some students presumably improved while many did not. An effective assessment program will tell us which educational sequences of coursework lead to the improvement of these learned abilities and which did not.[107]

---

[105] Ibid., 13; See: R.W. Farland and R. Cepeda, *Precollegiate Basic Skills in the California Community Colleges: A Report*, (Sacramento, CA: California Community Colleges, Office of the Chancellor, 1990), ERIC Document No. ED 317 256.

[106] Ibid., 15.

[107] Ibid., 17. Ratcliff's own work at the Center has been described in J.L. Ratcliff, *Development and testing of a cluster-analytic model for identifying coursework patterns associated with general learned abilities of college students: Final report, May 1990,* U.S. Department of Education, Office of Educational Research and

Ratcliff emphasizes the importance of such a reporting system in the political context; because "without such differentiation we are merely slinging mud at the institution,"[108] and yet most assessments and reporting efforts, he laments, have been at the institution level. Rather than ask the question which is usually implicit in most assessments—"Which colleges provide a better education?"—he believes that

> instead, the guiding question for assessment should be, "Which groups of students benefit most from which collegiate environments?" Only by answering the second question can we hope to show substantial increases in students' ability to reason critically, communicate clearly or to solve problems. A national assessment should identify between-state differences and between-student differences. The between-student differences could be defined according to academic ability or through focused study of groups with low participation or success rates in college.[109]

The lens through which individual students would be analyzed for their critical thinking abilities should focus on a developmental sequence of learning model. Ratcliff emphasizes that "we are not interested in determining the sum of all [critical thinking] learning experiences, nor the average performance of students in learning such abilities, but rather the effectiveness of the progression of learning in producing the desired results. . . It is the search for a more efficacious curriculum that leads us to assessment."[110]

Ratcliff and his colleagues at the National Center for Postsecondary Teaching, Learning and Assessment have been developing protocols to monitor selected courses, syllabi, and examinations, which could guide such a profile. Not only would these reports reveal how students were progressing, but they could begin to establish a constructive, non-threatening dialogue about which curricula sequences most enhanced critical thinking. Ultimately, he believes, states and institutions could feel properly invested in a national assessment program, as the focus would be less on institutional responsibility than on matching appropriate learning environments to the differing ability levels of students. This point is

---

Improvement, Research Division. Contract No. OERI-R-86-0016. (University Park, PA: Center for the Study of Higher Education, Pennsylvania State University, 1990).

[108] Ibid., 17.

[109] Ibid., 20.

[110] Ibid., 24.

relevant as a perspective on the present attempt to identify skills and abilities as described in Chapter 2.

> Rather than become embroiled in debate over what constitutes critical thinking or clear communications, we could begin an investigation of what existing measures overlap each other, which best describe student improvement, and which are most closely aligned with the curriculum of particular institutions or student groups. This second prong of investigation would move us closer to understanding *how* we may use assessment information to improve students' abilities in these key areas articulated in the NEGP objectives.[111]

Mike Knight puts the issue more boldly: "How will you convince the people on your campus that there is no bad news? If you can convince people of that fact," he predicts, "then you have resolved most of the major issues of organizational change." But such change involves the entire infrastructural environment, as composed of a number of people now getting paid predictable salaries. "How will you convince them? How will you create an environment where that is acceptable? Now if the belief is that if I reveal bad news I will suffer, the consequences are obvious."[112] Knight relates the history of a curriculum development cum assessment project at Kean College (NJ):

> When we began assessment on our campus, there was resistance. We anticipated resistance. We would have been astonished if there hadn't been resistance. This is the way we conceptualized this project. When we first discussed it, it was described as a student development project. Then, a curriculum development project. Then a faculty development project. And it is all of those, but it is more than that. Margaret Miller mentioned reward structure. It is an organizational change project. If it is not seen as an organizational change project, I do not believe it will be successful.[113]

Ed White, another author with an equally revolutionary approach to the national assessment of college student learning concurs with the thrust of Ratcliff's rationale:

---

[111] Ibid., 25.

[112] Michael Knight in Open Session.

[113] Ibid.

> the national assessment cannot be wholly top-down if it is to be effective. It must include the active participation of teachers, as well as the support of the local college community, if it is to lead to the improvements everyone seeks. Any assessment that is perceived to be external to the interests and values of a particular school will be an irrelevant annoyance rather than a spur to achievement in the school.[114]

The radical nature of Ratcliff's suggested system, however, caused the reviewers from the AAHE Assessment Forum to worry about its very premise—that faculty would embrace it as a positive complement to their own efforts. Several aspects of the system concerned them:

> (1) Can a campus-based system in which all the work of assessment is necessarily done by third-party experts (educational researchers), and that makes its case via elaborate computer-generated tables, be persuasive to a broad range of university faculty members? In current practice, the most successful assessment programs have started with faculty questions and had faculty themselves as the inquirer-data gatherers.
>
> (2) How persuasive to faculty advisors and to students are regression data, showing "tendencies," at best, from years ago, about a body of courses that is (one would hope) constantly changing and evolving?
>
> (3) Can data of this type from dozens of institutions in a state be cumulated in a way that is accurate, credible, and useful to state-level (not to mention national) decision makers? How could they interpret it? What, exactly, would it tell them? What decisions could appropriately follow from it? How would the general public react?[115]

"Effective writing cannot be examined in a test"

Venezky brought to the table the expertise of someone who sees assessment through the lens of literacy; ***Ed White's*** experience with teaching and analyzing writing convinces him that—when viewed as a higher order communication skill—writing "necessarily involve[s] critical thinking and problem solving."[116] Just as Ratcliff argued for a cumulative view of the progressive mastery of critical thinking skills, White emphasizes that effective writing entails "thoughtful or individualizing skills," and is really the culmination of a long learning process from letter formation, through spelling, vocabulary, grammar, sentence construction and so on—what he classifies as "imitative skills". . .

---

[114] Edward M. White, "Assessing Higher Order Thinking and Communication Skills in College Graduates: A Position Paper on Writing Assessment," (paper commissioned for the NACSL study design workshop by the U.S. Department of Education, National Center for Education Statistics, 1991): 20.

[115] Wright and Marchese on Ratcliff.

> We do not want original spelling or punctuation (though we often get them, at all levels) but we do want original thinking and independent critical problem-solving as part of higher order thinking skills. This difference between the imitative or socializing skills and the thoughtful or individualizing skills is crucial for thinking about assessment, for the imitative skills that may be appropriate to measure for younger children are likely to be inappropriate (or, at least, of minor importance) when we deal with higher order thinking skills.
>
> When we learn spelling and sentence structure, we do what we do specifically because that is the way they are *done*; that is, we do not think for ourselves. But when we develop arguments, conduct research, or solve problems, such imitation is not only insufficient, but it defeats the purpose; we must think for ourselves, as individuals, if we are to write well.[117]

The value of such skill is by no means tangential to Goal 5: "There is no question about the need for these individualizing skills as we consider writing as part of education for the workplace or for citizenship."[118] Echoing Banta and Nummedal, White emphasizes the importance of connecting the skills a national assessment of college student learning would target to the practical aspects of Goal 5. He cites the *SCANS* report, ("the problem solving that people do at work requires complex, situated learning"), and invokes the "citizenship aspect of Goal 5 as even more important:

> the basic premise of democracy demands citizens who do not merely do what they are told or think what they are told to think. Democratic political theory rests on the

---

[116] Ibid., 1.

[117] Ibid., 2.

[118] Ibid., 3.

premise that governmental authority depends on the "consent of the governed," and, as Thomas Jefferson argued, an uninformed consent is no consent at all.[119]

For White, the educational system needs to reflect this evolution in the appreciation of writing skills, and the way to do so, he asserts, is to regard writing not as a product, but rather as a process. Especially in the workplace context [emphasized by Banta and Nummedal], where White believes that both the value and the character of effective written communication becomes apparent:

> But the objection to defining writing as principally a product is nonetheless sound, for writing instruction and writing assessment. Writing on the job is more and more a matter of joint production, with teams of writers working together to produce reports, which are edited and published by a professional staff. The writer on the job must know how to produce drafts, and to revise work as a report develops.[120]

Such a focus, obviously, would be impossible to apply to an objective test, and difficult to apply even to an essay test situation. Thus, concludes White, the portfolio method has become the mode of choice for valid writing assessment.[121] "Higher education, in particular, has been most receptive to the concept. One estimate declares that fully one-third of American colleges and universities are experimenting with portfolio assessment for various purposes as we move through the 1990s."[122]

White thus strongly urges national assessment of college student learning developers to get in front of the curve:

---

[119] Ibid., 3.

[120] Ibid., 6.

[121] Ibid., 15 "Portfolio evaluation has recently been adopted from the fine arts to writing measurement. For example, the State University of New York at Stony Brook has been for some years using portfolios to measure upper-division writing proficiency by re-evaluating the writing produced in lower division composition courses; Alverno College in Minnesota bases its advising and curriculum on individual student portfolios maintained over the full span of the college years; and public school systems in New York City, Pittsburgh, and the state of Vermont have adopted portfolios for proficiency assessment." See K. De Witt, "In Vermont Schools, Test on How Well Students Think Draws New Interest." *The New York Times*, Sunday, September 1, 1991.

[122] Ibid., 16.

> A new national assessment should not merely repeat the most convenient and economical of present assessment procedures. The federal government should learn from what has gone wrong with many large-scale assessments in the past. In particular, any national assessment should be fully aware that reductive large-scale tests, given to students of widely different backgrounds and locations, have not supported improvement of instruction or learning. We should not administer an assessment that merely imitates the tests—and the mistakes—of the past.
>
> Assessment is much more than testing, and a national assessment worthy of the name will do much more than impose a national test on an unwilling nation. A true national assessment will foster and support regional assessments through portfolios and other process measures, in order to help students develop the high order thinking and problem-solving skills that the national interest demands. Through centralized funding, information and data gathering; by disseminating exemplary procedures and results; and by maintaining clear goals and standards, a national assessment can elicit from states and localities the participative assessments that lead to genuine and measurable improvement.
>
> We should expect improvement in performance as a result of the assessment, which must not be allowed merely to report what everyone knows is amiss. Such an assessment will require time, much consultation, and imaginative solutions to the problems of reliability and funding. But the power of this kind of assessment system to improve American education is immense. The present opportunity to turn assessment in this positive direction offers significant hope for the future.[123]

Lorenz Boehm of Oakton Community College (IL) concurs, and also favors portfolios as the better way to assess writing, which he believes to portray the very essence of critical thinking: "Writing is a perfect medium, perhaps *the* perfect medium, for assessing thinking. Assess the writing process, and you assess the thinking process; they are inseparable."[124] Boehm considers various current definitions of critical thinking to make his point, including, in Chapter 4, that of Robert Ennis. Here quoting Montclair State's Mark Weinstein—"Critical thinking in the disciplines (at the undergraduate level) requires mastery of the forms of inquiry. Embedded in language, such forms yield the tools for inventing, organizing, communicating, and utilizing the content of the various areas of human concern"—Boehm says that

---

[123] Ibid.. 19, 23.

[124] Boehm on White.

> Writing still works at both teaching and assessing critical thinking. All that changes is the number of faculty members using it; not the belief in its role. Says Toby Fulwiler: "writing progresses as an act of discovery; no other thinking process helps us so completely develop a line of inquiry or mode of thought." And the writing-as-process people, I among them, nod sagely, smile brightly, and think of portfolios. And rightly so. Given the range of writing/thinking a portfolio can accommodate, and given its ability to reflect the stages a piece of writing/thinking has gone through, it's hard to imagine a better assessment instrument.[125]

But Boehm is not overly sanguine the road to a national portfolio-based assessment system will be smoothly paved:

> Yet, my hunch is, in order to make that work, the national assessment administrative structure has to include a mechanism for bringing faculty and administrators into the fold; they too will need to see, believe in, and value the possibilities of the writing process. Yes, educators have to "recognize that some form of portfolio assessment. . is the most effective way to meet the need for national standards and a national assessment system" [as White says on p.25], but I don't believe we can't legislate that recognition. (We *can* legislate machine-scored, multiple-choice tests; they are more easily managed and are far less hassle.) In order for national assessment of critical thinking to be done through writing only, let alone a portfolio of writings, we are going to have to bring people to it. Make them believe, or else it will not work.[126]

White's forceful invocation of the possibility of reform was to become something of a flashpoint as the scholars gathered in Washington, where an interesting phenomena emerged. While many agreed with his insights about writing, and most applauded the zeal of his advocacy, ultimately his recommendation for a national portfolio system was not generally perceived as "do-able." [This term was a recurring motif of the discussion.] And while much of his thesis is couched in the history and analysis of writing, many insights and conclusions that he draws—especially about standards and bias, discussed in Chapter 3—

---

[125] Ibid.

[126] Ibid.

62

could serve national assessment of college student learning developers well, once translated into more specific questions about their task.

"Could Shangri-La, Wisconsin, be Exported to America?"

But an even greater example of the danger of throwing out the baby with the bath was the case of Alverno College. The prospect of converting, translating, and adapting the Alverno experience and the model that has emerged therefrom to the national assessment of college student learning presented—to some—an inconceivable odyssey. But the rigor of the papers presented by ***Marcia Mentkowski*** and ***Georgine Loacker***, which document a prodigious effort to adapt their experience at Alverno to the NCES debate, suggests that national assessment of college student learning thinkers need take a long, hard look at them. As ***Ted Marchese*** and ***Barbara Wright*** put it, in their review:

> An outstanding piece of work! We learned an enormous amount from this paper, which led us to a much deeper appreciation of the richness and complexity of the three abilities *and* convinced us of the importance of assessing the abilities in context.
>
> It now seems unthinkable to us to operationalize the three abilities strictly in terms of on-campus considerations; we need employers and other citizens at the table, all the more so because Goal 5 explicitly places these abilities in relation to citizenship and workforce development. The paper also makes a persuasive case for the need to assess the three abilities in use, and this may well mean looking at additional factors like interpersonal and "learning-to-learn" skills.
>
> To all that, we say Bravo! Let's keep this paper close at hand and remain open to its implications as we explicate the abilities and ponder their assessment. While the recommendations presented here are ambitious, they are also properly sensitive to the complexity of the three abilities and solidly backed by research as well as Alverno's extensive experience. By linking assessment of the three abilities to both accountability and improvement agendas, the paper suggests that one system, appropriately conceived, *can* serve both masters. If we follow this paper's lead, the effort to generate valid judgments will require significant investment—but at least we'll have some assurance that investment is serving educational ends.

Mentkowski and Loacker are both on the faculty at Alverno, a small women's college in Milwaukee, Wisconsin. Each was a primary author, and their 2/15 share of the original material was in actuality much larger, as each produced a distinct and—in the details of its exposition—comparatively stunning report for NCES.[127] Since 1973, the Alverno program has evolved into an ability-based curriculum with unique ways of approaching both undergraduate and postgraduate life, and this 18-year experience may provide the single most accomplished [and analyzed] assessment system in an American university. Loacker's paper argues "that a national assessment system should aim to achieve the dual purpose of improvement and accountability,"[128] while Mentkowski frames her analysis of the Alverno experience to display "what we have learned from our study of relationships between abilities/outcomes learned in college and those abilities as they are performed at work."[129] She suggests that, while her analysis stands alone, her colleague's paper provides "the context of the curriculum" at Alverno on which her analysis of the connections to the workplace are founded.

While significant elements of their discussions are taken up in Chapters 2 and 3, an overview is presented here, together with some reactions reviewers had to the papers. Alverno's core curriculum embodies a performance-based, outcomes-oriented approach where a graduating student must demonstrate "eight broad abilities: communication, analysis, problem solving, valuing in decision making, effective interaction, responsibility toward the global environment, effective citizenship, and aesthetic response at increasingly complex levels in a wide variety of settings and contexts."[130] The demonstration of these abilities is where the revolution in assessment lies, in that it is a continuous, dynamic, unifying element for the entire experience, both in college and after graduation. *Assessment at Alverno College* lists the principles basic to the process:

---

[127] Loacker, Georgine, "Designing a National Assessment System: Alverno's Institutional Perspective," and Mentkowski, Marcia, "Designing a National Assessment System: Assessing Abilities that Connect Education and Work," (papers commissioned for the NACSL study design workshop by the U.S. Department of Education, National Center for Education Statistics, 1991).

To reduce their material to a synthesis - which they thoughtfully provide themselves by way of summaries and charts - does not begin to do it justice. In sheer detail, as measured by computer bits, each paper was virtually twice the size of the next largest paper, produced by Paul and Nosich, which was itself a voluminously detailed exposition of critical thinking. Most of the other papers were little more than half the size of that. As Dick Larson put it in his review of Mentkowski:

"The paper and its appendices present so many conceptual frameworks and sub-frameworks, at high levels of abstraction and generalization, that one finishes the paper guessing that the Alverno people must have thought of all the possible concepts and interconnections of concepts one might imagine for describing critical thinking and problem solving as enacted in specific professional situations."

[128] Loacker, op. cit., i.

[129] Mentkowski, op. cit., i.

1. Assessment is an integral part of learning.
2. Assessment must involve a sample of behavior.
3. Assessment must involve a performance of an ability representing the expected learning outcomes of a course, a program, a department, and/or the institution.
4. Assessment involves expert judgment based on explicit criteria.
5. Assessment must incorporate structured feedback.
6. Assessment must occur in multiple modes and contexts.
7. Assessment must incorporate an external dimension.
8. Assessment is cumulative.
9. Assessment instruments must incorporate open-ended possibilities for demonstrating a given ability.
10. Self-assessment must be an essential part of assessment, as well as a goal of the process. It is an essential ability for the autonomous lifelong learner.

Loacker's own work elaborates on "how a faculty member and a group of faculty might use this generalized model:[131]

---

[130] See two basic documents produced by Alverno College Faculty, *Assessment at Alverno College*, Rev. ed., and *Liberal Learning at Alverno College*, (Milwaukee, WI: Alverno Productions, 1985).

[131] Loacker, op. cit., 3. See G. Loacker, L. Cromwell and K. O'Brien, "Assessment in higher education: To serve the learner," In Adelman, C., (Ed.) *Assessment in Higher Education: Issues and Contexts*, Report No. OR 86-301 (Washington, DC: U.S. Dept. of Education, 1986): 47-62.

> When any of us designs an assessment, we clarify what ability we are asking the student to demonstrate. We identify what *components* of the ability would be included in order to provide more focus for the design of the stimulus. Once we design the *stimulus*—whether it is a question or a set of directions, whether it will include something like a videotape—we determine more specific *criteria*. Then we use the stimulus with students and end up with a set of *performances*. We ask each of the *students to judge their performance* on the basis of the identified criteria.
>
> Then *we judge them* and *give feedback* that tells the students which criteria they met; which they showed deficiency in meeting, with evidence to clarify why and how; what they might have demonstrated that went beyond the criteria, and what they need to do further. Finally, our study of the student performances assists us to *evaluate the instrument and our own teaching in relation to it*. Did the stimulus work? Were the criteria clear and sufficient? Was there some aspect that I did not teach? That I did not give the students sufficient practice in?
>
> For every assessment faculty design, whether an individual one within a course or a more comprehensive one within the student's total academic program, they include all of these elements even though they might not always work with them in the same order.[132]

Loacker's analysis is based on ten principles that the Alverno experience has taught those who have participated and analyzed it. While some of these recur elsewhere in the discussion in particular contexts, they provide a convenient key to her paper:

1. An ability-based performance assessment system, with certain key elements,[133] can work both to evaluate student performance and to develop student knowledge and ability.

2. Making expected outcomes explicit and public to all, identifying developmental criteria for performance, and communicating them to students ahead of time,

---

[132] Ibid., 3.

[133] These are: Public abilities/outcomes and criteria, multiplicity of performances across varied contexts, expert judgment, feedback, and self-assessment.

contributes to effective performance by making learning more accessible and enabling performance.

3. Feedback on performance in relation to developmental performance criteria and the opportunity to interpret that information leads to further learning and improvement of student and program performance.

4. Students learn complex abilities, including self-sustained learning, in the curriculum through a variety of contexts.

5. Students can transfer abilities when they are assessed in contexts that are valid for what students learned and for how they will perform abilities later.

6. When an assessment system examines changes in student abilities/outcomes over time, including who changes and why, and relates those changes to the curriculum, the system yields information necessary for meaningful improvement.

7. We can validate an ability-based performance assessment process, and institute an instrument validation process that gradually improves instrument validity. We can establish the educational value, impact, validity and effectiveness of the abilities/outcomes.

8. A dynamic assessment system incorporating input from and feedback to faculty, as well as administrators, provides for the effective use of information to keep abilities, performance criteria and standards responsive to and in advance of the needs of our society.

9. Creating a context for assessment is as important as creating the assessment method.

10. The effectiveness of an assessment system concerned with the improvement of learning depends partially on a coherence that comes from the following articulated components:

    - educational values, assumptions and principles that are tied to the mission statement of the institution,

- an assessment theory (what are the components of good assessment?) consistent with those values and assumptions,
- a psychometric theory (how do we best measure and credential performance and give feedback to students on their abilities?) consistent with those values and assumptions.[134]

Loacker's colleague Mentkowski builds on this analysis of Alverno's system to focus her study on the aspect of Goal 5 that the Alverno philosophy inherently addresses—the connection between assessing abilities and workplace relevance. Alverno is not only extremely experienced at assessing its students and graduates, but is continually refining the focus of such assessments to assure that the skills being developed are those that will work best at work. Thus—even though many commentators and scholars felt it virtually inconceivable that the national assessment of college student learning could successfully scale these twin peaks of Goal 5—the Alverno system's essential structure operates from the basic premise that assessment must be a lens to view—not just abstract and possibly irrelevant abilities, but rather—the overall suitability of women for effective and successful postgraduate lives.

It is almost as if the governors and the president had conducted a tour of Alverno, and then emerged to dream and aspire to the national goals, at least to Goal 5. The question seems to be—not whether the Alverno experience provides a valid, extraordinary and felicitous exemplar; but rather—whether the system developed there could ever be extrapolated to the nation at large. While this issue was not settled at the NCES workshop, [nor will it be here] there is another imperative that seems to emerge from the Alverno papers. Because the system is so elaborately structured with rich elements born of successful experience, it would seem incumbent on national assessment of college student learning developers to consider first the substance and the many rich particularities of the Alverno model, before concluding that context is too much of a special case to be adapted to a national scale. Even if this turns out to be so, the serious consideration of how the system might transfer should reveal a number of vital issues which have been tested under invaluable, real-world conditions.

In this spirit (though Mentkowski's overall thesis rests upon a number of assumptions, and leads to seven principles, that many may feel go further than the national assessment of college student learning should or can go) the view of assessment she portrays can hardly be ignored by national assessment of college student learning developers. That is to say, discrediting her assumptions as *not rigorously proven or unrealistic* doesn't nullify the potential of an Alverno-like assessment system, but instead only reveals the enormous scope of the challenges faced by national assessment of college student learning developers. Most of her assumptions do not, in fact, underpin the validity of the system so much as they describe an ideal context in which assessment is postulated to be most effective:

---

[134] Ibid., 7.

1. We assume that the context for relating education and work includes before, during and after college.

   College population demographics show that the majority of college students no longer consist of 18-year-olds who begin college directly from high school, complete their education in four years, and then either go on to graduate school or enter the workforce. Rather, college students are likely to be from a broad age range, with a heterogeneous work experience. Many combine college with full-time or part-time employment; our own students work before, during and after college.

2. As workplace demands become more complex, employers and the professions are putting more and more resources into education, training and continuing education programs.

   Educators in corporations and the professions are expecting more support in this effort from colleges and universities. Thus, more persons will be coming in to or back to college across their lifespan. With so many college students working before during and after college, and so many workers requiring more learning opportunities, we need to see college learning in the context of work, and work in the context of college and other kinds of work-related learning. While this paper often uses language implying that transfer "from college to work" is the only interest, we assume that work influences college as much as college influences work.

3. College and university students and faculty, and employers and employees in business, community organizations and the professions, are joined together in a mutual enterprise.

   All these persons are invested to some degree in improving learning opportunities at college and in the workplace. This paper assumes that faculty and employers, students and employees, have a common interest in assessing abilities that prepare one for twenty-first century demands. These include participating in a global economy and global citizenship. The assumption is that a national assessment system can call on and count on this common interest.

4. Workplace and citizenship demands are becoming so complex that, contrary to past perceptions, college is now a place where students are learning abilities that will be important for day-to-day work as well as for personal or leadership goals.

   The kinds of complex thinking, communication and problem-solving skills learned in college are in great demand in the workplace as well as one's personal life. Interpersonal abilities, leadership qualities and self-directed learning skills are essential across all of these situations, particularly as business and industry moves to more of a "total quality management" environment.[135] We assume that meeting the needs of the workplace and citizenship demands means continued learning across the lifespan.

5. Both college faculty and employers have a joint interest in individuals meeting expectations for citizenship, and they assume that abilities developed in college and performed at work are ones that transfer to service and citizenship roles.

   Information from assessment can assist both faculty and employers to determine the degree to which graduates and employees are meeting their own expectations, those of faculty and employers, and of those roles outside of work. These abilities should serve the individual not only at work, but also in roles outside of work: in personal life, service and citizenship.

6. Any national assessment system will consider that abilities, assessed in context, will change, because the demands for performance and the contexts in which abilities are performed, will change.

---

[135] K. DeWitt, "In Vermont schools, test on how well students think draws interest," *New York Times*, Education Life, August, 1991.

7. We assume that readers accept the idea of a national assessment system rather than a single national test.[136]

Like Loacker, Mentkowski's rich analysis is replete with supporting data and references, and follows the same outline of exposition: a series of principles learned, each of which leads to specific recommendations for the national assessment of college student learning, as well as suggesting associated implications, issues and questions:

1. Abilities can be defined in ways that connect education and work.

2. Abilities defined with multiple components, and as integrated, developmental and transferable, are likely to make sense both to educators and employers. These can be assessed within professional roles in appropriate contexts.

3. Thinking critically, communicating effectively and solving problems are abilities common to college education and work; effective performance at work is integrated; it is made up both of intellectual and interpersonal abilities.

4. To effectively transfer college-learned abilities, students need to develop learning to learn skills, or self-sustained learning. Assessment that incorporates feedback and opportunities to self-assess, fosters self-sustained learning.

5. Comparing faculty-defined abilities to those demonstrated by outstanding professionals enables faculty to identify abilities students need for particular professions.

6. Complex abilities that connect education and work—including self-sustained learning—can be assessed in graduates' work performance in a variety of professional contexts. Some abilities can be linked to college learning, and some distinguish effective performance at work.

7. Faculty, professionals, and employers will invest in understanding the relationship between education and work if they can create contextually rich descriptions of performance in relation to their judgment of what abilities to develop.[137]

---

[136] Ibid., 1.

[137] Ibid., 3.

Most participants at the workshop were aware of the Alverno model, though the Mentkowski and Loacker papers—framed as they were to fit the NCES outline—placed the issue of its suitability stage front. Many of the concerns expressed about Alverno can be seen as entirely consistent with Mentkowski's paper, except that the assumptions she bases analysis upon were *not*, many felt, widely shared, however laudable they may be as platonic ideals. Peg Miller from the State Department of Education in Virginia said the Alverno program "has been an inspiration to those of us who have struggled to institute state-wide assessment programs. . .".

"It has not, however, apparently been a practicable model for the institutions, at least in Virginia."[138] While she doesn't puzzle too much about its failure to catch on at "the large, complex universities in which undergraduate education is only one of a number of sometimes conflicting priorities," she does wonder why the small liberal arts or two-year colleges whose "missions focus primarily or even exclusively on educating the undergraduate student" have not been able to benefit from the Alverno example. Another reviewer on Loacker, ***Mary L. Tenopyr***, concurs: "There are some severe questions about whether an assessment system tied to the mission of a particular institution can be expanded into a national system. Alverno's program might be extended and amplified so that it can serve as a basic model for liberal arts institutions, but how will it serve full universities or more technically oriented institutions like the California Institute of Technology?"[139]

The most apparent stumbling block, as was the case with Morante's GIS story in New Jersey, was faculty. Said Miller:

> It seems that for a faculty to organize its pedagogy and curriculum around assessment as Alverno has done involves a change in faculty culture that is quite profound, and in Virginia at any rate, faculty have by and large resisted the transformation.[140]

Whether the suggestions offered by Lenth and Ratcliff offer a structural way to finesse this problem by involving faculty in a more fundamental way in the process, and by structuring the assessment in ways that reinforce rather than threaten a faculty's hegemony, remains an open question.

---

[138] Miller on Loacker.

[139] Tenopyr on Loacker.

[140] Miller on Loacker.

In addition to this apparent unwillingness of an institution to "link assessment to serious curricular and pedagogical change," Miller warns of other foreseeable problems in an Alverno-like national assessment of college student learning. Coherence is a real stumbling block, she believes, suggesting that there may be some critical mass of members in an assessment network system beyond which principles cannot be agreed upon. Noting that "Alverno has had some success in getting three institutions to agree to 'principles of ability-based performance assessment,' and that another group of 11 is "currently synthesizing educational assumptions that are common across their institutions," Miller reports that experience in Virginia suggests that educators have a tendency to approach assessment in idiosyncratic and personalized ways, and will often clash over the issue of "transfer articulation." One shouldn't dismiss these local views as short-sighted, however. Miller believes that

> campus-based assessment is the kind most likely to support rather than damage the teaching-learning relationship, acknowledge what has been a fruitful diversity of institutional purpose in American higher education, honor faculty control over educational matters, and do some good for the individual student.[141]

But, she continues, such systems are "costly, redundant, slow-moving, and unlikely to produce easily understood results." More power to the institutions where such systems have been established, she says, but let's not deceive ourselves about the political reality:

> The national assessment movement, much as we may wish it were otherwise, seems to be primarily driven by a desire for accountability and hence a desire for relatively simple and comparable data about what American students know and are able to do at any point in time as compared to the previous year or decade. In this context, questions of institutional mission and practices are secondary, and an assessment system characterized by complexity, multiplicity, and a lack of stability is not going to fill the bill.
>
> It seems to me that any national assessment system should preferably build on the good assessment being done on some campuses now or at least not damage or supplant those assessment programs. But it cannot rely on them, since they are far from ubiquitous and not particularly well suited to the job the national assessment system is

---

[141] Ibid.

> supposed to do, give a coherent and probably simplistic picture of higher education's results and progress.[142]

Practicality was a theme also stressed by another Alverno reviewer, Richard Larson, from the Lehman College of the City of New York, who reflects a "strict constructionist" view of Goal 5 when he asks "I am not sure I understand our assessment of students' communicating effectively, critical thinking, and problem-solving abilities might best occur sometime *after* the student has completed undergraduate work, or that it might best be carried on in the context of the workplace."[143]

> To be sure, there is no reason why on-the-job experience should not be part of what a graduate brings to the assessment, but if such experience is going to be included in the preparation to be assessed, how far beyond college experience do we go? And if we go beyond college experience, are we learning how well we in higher education are achieving, generally, our educational goals, or something else (e. g., how fortunate the students are in getting positions that offer good workplace training)?[144]

Furthermore, he asks, even if there were a clear rationale and target for workplace assessment of graduates, the practical obstacles are formidable, including tracking students into the workplace, trying to adjust for differing contexts and settings where the subsequent assessment might take place, and the training of so many investigator/observers. Asks Larson:

> Given what Dr. Lenth has taught us about the wide variation in interest in assessment from state to state, could we hope to get any agreement about educational leaders, public and private, in various states, about the value of carrying on this complex process? And, given what Dr. Morante has taught us, could we be sure that an initial agreement, however hard won, would survive shifts in political winds from state to state and in the nation?[145]

---

[142] Ibid.

[143] Larson on Mentkowski.

[144] Ibid.

[145] Ibid.

Finally, Larson poses a number of questions about the details of the system proposed by Mentkowski:

> How does the assessment procedure described by Dr. Mentkowski work, anyway? What kinds of behavior are selected for examination? How are they selected? What exactly is the procedure for determining the characteristics or quality of the performance given? Who makes those decisions and how are these people trained to make them? (In a national assessment, we'd need lots of interpreters and lots of people to train the interpreters.) What exactly is *looked at* in the performance, against what criteria or evaluative grid?
>
> There are, it would appear, different kinds of interviews and strategies for interviewing, and I gather that there would need to be, in a large-scale study, some agreement about the kinds of interviews to be undertaken. And, if we need to do so, can we make the appraisals of the individuals studied comparable enough so that we can reach some conclusions? Can we train the people who will conduct the interviews and appraise the performance?[146]

The interview process forms such an integral part of the model that ***Ronald G. Swanson***, another Alverno reviewer, felt it important to note:

> One of my biggest concerns about applying the Alverno model nationally is its dependence on lengthy, complex interviews. First, as has been pointed out often in the past, interview data can be called into question for many reasons—lack of standardized administration, variable judgments about what is to be recorded and what is not, the subjectivity of data obtained orally in a face to face setting—to name a few. Even if some of those shortcomings could be dealt with (and some of them can), how in the world could the interview technique be applied on a national scale?
>
> I was also a bit concerned about using experiential reports obtained during interviews as a basis for determining how students learn. I guess that all of my previous experience with the rigors of the scientific method leaves me less than enthusiastic about building a body of data based upon subjective reporting.

---

[146] Ibid.

Thus, while these and other specific criticisms of the Alverno model must be appraised and evaluated, they do not seem to undermine the possible value of this rich source of actual assessment experience in America, as the national assessment of college student learning begins to take shape.

A Final Caveat

Stephen Dunbar made the case for inaugurating systematic research to develop a national assessment of college student learning in general, and the component abilities in particular.

> Because *America 2000* marches in a direction that involves some of the most controversial uses of educational measurement in content domains where much uncertainty exists about construct definitions, NCES should proceed slowly with the development and implementation of an assessment program for the national assessment of college student learning. This recommendation is not a mere attempt to rain on the *America 2000* parade. The state of knowledge regarding the measurement of higher-order skills at the college level, not to mention the mapping of national achievement scales onto social utility scales, is sufficiently limited that no procedure can be recommended for a national assessment of college students because of the good data that it has produced to date.
>
> Instead, a program of systematic research directed at the development of procedures and instruments for national assessment of college student learning should be initiated. A carefully designed research program, predicated on the eventual implementation of a federal system of postsecondary measures of achievement, can itself provide preliminary indicators of activities in college classrooms that enhance critical thinking, communication, and problem solving.[147] Dunbar urges a closer look at Goal 5, in the context of both what is feasible and what is socially desirable.

"*What to measure*," he insists, "is not transparent from the language of the Goal 5 objective."[148] Is it possible to reach a consensus on what problem solving, critical thinking and communicating effectively consist in? Yes, he thinks, and the way to proceed is

---

[147] Dunbar, op. cit., 17.

[148] Ibid., 3.

"traditional instrument development procedures that entail the delineation of content domains, their translation into tables of specifications, and the writing of performance tasks and test questions."[149] Anything even approaching Dunbar's "traditional" and very coherent approach has yet to be undertaken, though it may come about within the sequence of RFPs. Notwithstanding his caveat, most of the authors did begin to approach the identification of the skills themselves, and these first steps are discussed in Chapter 2, "Which Skills Should Be Assessed?"

---

[149] Ibid., 5.

2. WHICH SKILLS SHOULD BE ASSESSED?

Objective 5 from Goal Five, as specified by the National Education Goals Panel (NEGP) in *America 2000* is more a signpost than a map:

> The proportion of college graduates who demonstrate an advanced ability to *think critically, communicate effectively, and solve problems*, will increase substantially.

Objective 5 does not instruct policymakers and educators in any detail as to *how* this increase will be achieved. By delineating the categories where improvement is desired, however, the framers did lay the groundwork for measurement. "Increase" suggests quantifiable results, which could be tracked over time. Thus NCES asked its assembled scholars and advisors to suggest how to approach the question of quantifying performance of college graduates on what have been called the "Objective 5 abilities," i.e., critical thinking, effective communications, and problem solving. If there were some thesaurus or manual to consult for the details of what it is that constitutes communicating effectively, problem solving, and critical thinking, then addressing the issue in this chapter would be redundant. No such consensus seems to exist, however, and two other difficulties further complicate the task of identifying a list of skills and abilities.

First, even if there were an agreed-upon body of skills that constituted critical thinking, the national assessment of college student learning discussion would still revolve around the search for a coherent body of *assessable* skills. Second, a problem arises in the semantics of the Goal as sought through its objectives. Goal Five:

> By the year 2000, every adult American will be literate and will possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship.

Even though Objectives One through Four relate to the Goal, the quantification called for in those objectives is either enormously general ("Every major American business will . . .") or straightforwardly measured ("an increase in the number of programs", proportion of a target group graduating, etc.). Thus a number of advisors in the NCES exercise felt compelled to consider the larger issues of "global competitiveness" and "responsibilities of citizenship" as part of their charge from NCES, and so the seemingly straightforward charge to identify the Objective 5 skills and abilities revealed a wealth of complexities and considerations implicit in the totality of Goal Five.

Don Lazere, for example, from California Polytechnical Institute at San Luis Obispo, was concerned about what he perceived as a lack of emphasis

> throughout these meetings. . .concerning every adult American having the knowledge and skills necessary to exercise the responsibilities of citizenship. I teach English Literature and Composition, so I'm not a political scientist. But I'm constantly overwhelmed in all of my writing and literature courses by the fact that whenever issues of civics, citizenship issues, come up the appalling level of student ignorance and indifference toward citizenship.
>
> So I would like to urge here that, in the future activities and projects of this project, that there be a strong emphasis on the application of critical thinking, communicating effectively, and problem solving to the development of the rights and responsibilities of citizenship, and that when aspects and criteria for critical thinking and so forth are defined, that there be a section defining and applying them to exactly what rights and responsibilities of citizenship need to be highlighted in reference to critical thinking, communicating effectively, and problem solving. Maybe some political scientists might be brought into this effort, along the way, at that stage.[150]

While debate at the workshop was to reveal that most advisors were aware of the option (and the theoretical value) of addressing these larger questions of citizenship and the workplace, the pragmatic approach—which was limited to what could be done in a fairly straightforward fashion—drew much support. As Swanson said in his commentary on Mentkowski:

> While the fifth goal does talk about competing in a global economy and exercising the responsibilities of citizenship, the fifth objective is itself much more limited in scope. Figuring out a way to accomplish the fifth objective will be a Herculean task without attempting to directly assess abilities that connect education and work. I believe it was to be assumed that an improvement in higher order thinking skills would de facto enable graduates to perform better on the job. Please understand that I am not saying that we should not attempt to directly link education with work. What I am saying is

---

[150] Donald Lazere in Open Session.

> that such an effort is, in my opinion, beyond our charge and perhaps not possible in the time frame available. My plea here is for realism.[151]

Swanson's view was echoed often, by other authors and reviewers, and repeatedly during the workshop. When he concedes, however, that "I am not saying we should not attempt to directly link education with work" [and, as others said, with citizenship], he is acknowledging that even a strict effort to catalog effective communication, problem solving, and critical thinking for the purposes of assessment will inevitably invoke the *contexts* of workplace suitability and effective citizenship. Thus, while many objected to Alverno's emphasis on linkage as unrealistic—in a national assessment of college student learning that would be "do-able" in the near term—few objected to its underlying principles. The question isn't whether citizenship preparation and workplace relevance can be factored into a national assessment of college student learning, but rather *how* they factor in. Most agreed they were not explicit and distinct skills, but rather the effect of mastering and applying many other skills. Thus were these twin lenses, if you will, often fitted to the microscope through which scholars were analyzing and dissecting possible lists and catalogs of skills and abilities.

The discussion about a manageable, do-able national assessment of college student learning wasn't limited to narrowing the assessment to a confined list of Objective 5 abilities. Marchese flatly stated, citing six years of American experience in assessment by faculty trying to relate teaching and learning, that "we don't know how to answer the public's question[s] embodied in Goal Objective 5, [which] is: "What is your contribution to student learning? With respect to these three abilities, what do your graduates know, and can they do what your degrees imply? We can't answer the public's questions"[152] about that.

> We also feel that the time that we're going to have to answer these questions is not ten but perhaps two years, maybe three. That we're never going to have a hundred million dollars, we might have one or three million. And that we need to do the best focused thing in the time immediately ahead to teach ourselves how to take the experience and the knowledge that we already have, and devise ways of answering the public's questions about our contributions to student learning.[153]

---

[151] Swanson on Mentkowski.

[152] Ted Marchese in Open Session.

[153] Ibid.

To the specific charge of identifying target Objective 5 skills and abilities Richard Paul and Gerald Nosich bring their broad and deep background in the critical thinking movement, and their experiences as director and assistant director of the Center for Critical Thinking at Sonoma State University (CA).

But they do *not* have one coherent instrument to recommend because, they contend, even the latest and more advanced models of assessment fail to capture the richness of this body of thought, "principally because the [critical thinking] concept has been developed extensively only over the last ten years, and therefore has not had time to permeate already developed assessment tools."[154] By contrast, ***Peter Facione*** is more optimistic: "There is reasonable accord about an appropriate and rich conceptualization of critical thinking. We know how to conduct valid and reliable critical thinking assessment, and have developed at least a few instruments suitable for such an analysis.[155]

Thus, a fundamental question is posed: Can an assessment that adequately captures the target effective communicating, problem solving, and critical thinking [skills] be created in a fairly straightforward fashion, once a consensus on these skills is more clearly articulated? Or, conversely, do these abilities suggest such a revolutionary view of assessment as to require a massive effort that would entail extensive research and development, and perhaps many years to complete? This conundrum was constantly raised in reviews and workshop discussion.

Though almost all of the NCES advisors are working (and publishing) academic scholars, many of them are also battle-scarred veterans of the continuing struggle in American education to implement the best viable system, which struggle (by definition) continues even as more research and interim feedback, results, and scholarly evaluation indicate ways in which that system can and should be improved. In the political world where systems and programs get funded, the response to a call for reform may often precede definitive academic consensus about how best to implement that reform. Thus it is, to some extent, with the call for a national assessment of college student learning that would be based on a current best reading of the principles of critical thinking. From the NCES workshop deliberations on the question of skills and abilities emerged three different ways to approach the question: What should be the target? Practically speaking, what is a realistic target? And what sort of research might begin the process of bridging the gulf between the two?

---

[154] Paul and Nosich, op. cit., 2.

[155] Facione on Nummedal. The objective critical thinking assessment tool referred to is the newly published "California Critical Thinking Skills Test: College Level," 1990, The California Academic Press, Millbrae, CA. 94030.

The Critical Thinking Skills

The Search for Definition and Consensus

Trudy Banta was among those calling forcefully for a panel to be established in order to frame and clarify the task of identifying the skills and abilities to be assessed. She bases her recommendation on years of experience: "Since 1982, I have coordinated a comprehensive student outcomes assessment program at the University of Tennessee, Knoxville (UTK). In terms of its longevity, the extent of participation by units within the institution, numbers of students tested, and comprehensiveness of its ongoing assessment related research agenda, the outcomes assessment program at UTK is unique among those at U.S. research universities."[156] She notes that "no effort has yet been made to develop a broad national consensus among faculty regarding definitions of critical thinking and communicating, much less about ways to teach these concepts."[157] She reveals and examines five implicit assumptions behind Goal 5—one of which is that "The abilities can be defined and agreed upon"[158]—and would set the task of defining the abilities as one of the panel's foremost. She cites Facione and the American Philosophical Association Delphi process[159] in describing the obstacles to a consensus definition as "enormous," but believes that "building that consensus is absolutely essential."[160]

---

[156] Banta, op. cit., 2.

[157] Ibid., 2. The recommendations for further research by Banta and a handful of other advisors will be taken up later in this section.

[158] Ibid., 3.

[159] Peter A. Facione, The Delphi Report, Prepared for the American Philosophical Association: "Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction," (Millbrae, CA: California Academic Press), 1990. Eric # ED 315 423. A 22 page "Executive Summary," which includes all tables, findings and recommendations of the Delphi Panel, is published by the California Academic Press, 217 La Cruz, Millbrae, CA. 94030.

[160] Banta, op. cit., 5.

Two reviewers from the AAHE Assessment Forum explained why they thought it premature to collapse the inquiry into one solely limited by a search for consensus and practicality. Barbara Wright and Ted Marchese on Banta:

> We wonder whether consensus on a single definition of critical thinking et al. is really "essential" or even desirable, much less possible. This may be the conventional wisdom when we're looking at a traditional high-stakes testing situation. But moving away from that context, it can be argued that such consensus would lead to a disastrous reductionism, a dangerous impoverishment of what we mean by "critical thinking." Doesn't such an assumption impose a kind of scientific rationalism on the chaotic richness of human life, intellectual styles, and contexts for thought? Don't we thus confuse "uniformity" with "quality"?
>
> Of course, diversity doesn't guarantee quality, any more than uniformity does. But [analogous] to the value we place on biological diversity, in the interests of robustness, adaptability, and fairness, it makes sense to encourage or at least accommodate the widest possible range of variation in intellectual processes. The participants in this gathering can doubtless think of many ways to handle the issue of definition to allow maximum flexibility.[161]

Adds Norman Frederiksen, formerly with the Educational Testing Service (ETS): "I agree. There is too much variability among deans and professors in different kinds of colleges and universities to expect anything like a consensus on goals and objectives."[162]

In referring to communicating effectively, critical thinking, and problem solving as the "Objective 5 abilities," Banta is less controversial when she agrees "with Cuban[163] and others, however, that critical thinking, reasoning, and problem solving are virtually indistinguishable,"[164] and should be collapsed into the term—and the concept—critical thinking. As the NCES workshop exercise and deliberations ensued—though there were occasional comments made about the subtleties of problem solving and effective communication being slighted—most of the advisors and commentators proceeded by global reference to critical thinking, without too much quibbling about whether problem solving and communicating effectively introduce wholly distinct domains. [A notable exception in his paper and at the workshop was Ed White, whose forceful advocacy of a modified portfolio system to evaluate writing is taken up later in Chapter 2 and also Chapter 3.]

---

[161] Wright and Marchese on Banta.

[162] Frederiksen on Banta.

[163] L. Cuban, "Policy and Research Dilemmas in the Teaching of Reasoning: Unplanned Designs," in *Review of Educational Research*, 54(1984): 655-681.

[164] Banta, op. cit., 7.

Holding a brief for maintaining the distinctions were a minority of scholars, one of whom, ***John Chaffee***, proffered the program developed at LaGuardia Community College at The City University of New York as an example of the power of the problem solving category. One of the three components of the program there is described as "Reasoning and Problem-Solving." Chaffee quotes an evaluation from ETS which concluded that "the program fosters the development of students' thinking abilities at both general and specific levels. At the general level, teachers perceive more respect for the thinking process, more tendency to bring a "habit of thinking" to their classes. At the specific level, teachers reported instances of transfer of such skills as breaking problems into parts, classifying, organization of thought, asking questions, separating facts from opinions, and assessing points of view."[165]

Thus a major question that was raised, but not answered, was whether problem solving per se needs to be approached with a separate view. Paul says: No! referring to the "rich conceptualization" of critical thinking he recommends as embracing all such elements. . . .

> Now when you consider these richly—and here I would make an observation based on my own thinking, which you may or may not agree with—that they tend to converge. So when you try with a rich concept of critical thinking to distinguish it from effective problem solving, you have great difficulty. Because if you've got somebody you call a very good critical thinker who's not very effective in solving problems, you have a virtual contradiction in terms. . . .
>
> So, it seems to me you want to be very careful to bring in those people in the areas, who approach the areas, richly and broadly, with a sense of inter-disciplinariness, and not those who speak for the area in a very narrow, specialized way. And I think this

---

[165] Amendment to Chaffee on Nummedal.

> is a very important thing which will bear on the credibility, and the usefulness, of the assessment instrument that emerges.[166]

. . .Whereas Banta would probably say: Yes! Maintain the distinction in order to emphasize elements of thinking and problem analysis that draw on socializing skills and which require the thinker to cooperate with others and create a kind of gestalt in which to frame the problem before the more particular, analytical critical thinking skills come into play. ***Mark Weinstein*** from Montclair State (NJ) is another major figure on the American critical thinking scene, who

> recommend[s] that not only should people like engineers engaged in problem solving have their say at what should be done, but people who accept the universal areas of concern identified with critical thinking—and maybe even the universal dispositions of mind that aid and abet critical thinking—people who have this legitimate concern but see that concern articulated through specific areas of study (natural sciences, social sciences, humanities, professional studies especially) be invited to report on what critical thinking these generic, universal skills look like when identified, articulated, manifested, and assessed within these special areas of concern.[167]

While pragmatic compromise on an assessment process that could be implemented quickly may be unavoidable, the critical thinking experts at this early stage of deliberating the national assessment of college student learning clearly saw their mission: to deliver a substantive and forceful message about what critical thinking actually *is*. Facione provides a list of resources,[168] and Paul and Nosich urge that the national assessment of college student learning development process fully consider the elaborate catalog of skills and associated implications that has been developed in the critical thinking school, concluding that "If anything less than this concept and its central aspects is assessed, the ultimate goal of fostering higher order thinking as an academic, social, and vocational need will be ill served."[169]

---

[166] Richard Paul in Open Session.

[167] Mark Weinstein in Open Session.

[168] Facione on Nummedal. "To become better connected with advances in theory development and empirical research in the area of CT, one might also contact The Center for Critical Thinking at Montclair State College in New Jersey, directed by Mark Weinstein; The Center for Critical Thinking and Moral Critique at Sonoma State University, directed by Richard Paul; and the Institute for Educational Research and Development at the University of Newfoundland, particularly Dr. Stephen Norris. Other persons with practical experience and technical, scholarly expertise in CT assessment who should be consulted include Joanne Carter-Wells, Dept. of Reading, CSU, Fullerton, and Barbara M. Lawrence, Coordinator of Student Outcomes Assessment at Idaho State University, and Marcia Mentkowski, of Assessment Office of Alverno College."

The National Council for Excellence in Critical Thinking Instruction has developed a working definition of critical thinking, which the authors say captures the basic idea common to practitioners and researchers in critical thinking:

> Critical thinking is the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action.[170]

A fundamental premise of all learning, they stress, is the need "to reason to all basic conclusions and solutions, and to reason through and across the curriculum."[171]

The Facione Delphi report cited by Banta and published in 1990 identified a "Consensus List of critical thinking Cognitive Skills and Sub Skills," as follows:

(1) Interpretation (Categorization, Decoding significance, and Clarifying meaning),

(2) Analysis (Examining ideas, Identifying arguments, Analyzing arguments),

(3) Evaluation (Assessing claims, Assessing arguments),

(4) Inference (Querying evidence, Conjecturing alternatives, Drawing conclusions),

(5) Explanation (Stating results, Justifying procedures, Presenting arguments), and

(6) Self-Regulation (Self-examination, Self-correction).

---

[169] Paul and Nosich, op. cit. 2.

[170] Ibid, 2.

[171] Ibid., 2.

Facione describes the effort that produced this catalog:

> In 1990, a national panel of 46 critical thinking experts, drawn from several different disciplines and kinds of colleges, completed two years of careful collaborative work with the publication of a well-argued, detailed consensus regarding the core college level critical thinking skills and dispositions. Each skill and sub-skill they identified qualifies not as a discipline specific factor, but a genuinely transferable cognitive skill that can be used in either a social or individual real life problem-solving context.[172]

Facione characterizes the National Council for Critical Thinking's draft as reinforcing and confirming the Delphi report.[173] In their NCES paper, Paul and Nosich outline this catalog of abilities, and provide numerous examples in the context of sample questions. They state 21 objectives of a process to assess higher order thinking, which they intend to offer as a summary of how critical thinking expertise could be applied to a national assessment of college student learning. Those objectives which seem to go to the selection of skills directly:

> (1) [The process] should assess students' skills and abilities in analyzing, synthesizing, applying, and evaluating information.[174]
>
> (2) It should include items that assess two fundamental skills: that of thoughtfully choosing the most reasonable answer to a problem from among a pre-selected set; and also the skill of formulating the problem itself and of making the initial selection among relevant alternatives.

---

[172] Facione on Nummedal, 3.

[173] Facione on Paul and Nosich, 3.

[174] Facione commends Paul as the "philosophical guru whose energizing vision" produced the Center for Critical Thinking and a wealth of valuable CT insights and literature. Nonetheless, he feels compelled to respond to the details of the catalog. He says that "the positive value of this proposed criterion [#1] is to point us toward *content validity*, a vital component of any sound assessment design. . . Unfortunately Dr. Paul's way of putting criterion #1 compresses the theoretical concern for content validity with a partial list of some CT skills. A well-formulated criterion would separate the theoretical consideration (content validity) from an incomplete analysis of that content." [2].

(3) It should concentrate on thinking skills that can be employed with maximum flexibility, in a wide variety of disciplines, situations and contexts.

(4) It should concentrate on assessing the fundamental cognitive structures of communication at the college-level, for example:

with reading or listening, the ability to

(a) create an accurate interpretation,
(b) assess the author's or speaker's purpose,
(c) accurately identify the question-at-issue or problem being discussed,
(d) accurately identify basic concepts at the heart of what is said or written,
(e) see significant implications of the advocated position,
(f) identify, understand, and evaluate the assumptions underlying someone's position,
(g) recognize evidence, argument, inference (or their lack) in oral and written presentations,
(h) reasonably assess the credibility of an author or speaker,
(i) accurately grasp the point of view of the author or speaker,
(j) empathetically reason within the point of view of the author or speaker.

with writing and speaking, the ability to

(k) identify and explicate one's own point of view and its implications,
(l) be clear about and communicate clearly, in either spoken or written form, the problem one is addressing,
(m) be clear about what one is assuming, presupposing, or taking for granted,
(n) present one's position precisely, accurately, completely, and give relevant, logical, and fair arguments for it,
(o) cite relevant evidence and experiences to support one's position,
(p) see, formulate and take account of alternative positions and opposing points of view, recognizing and evaluating evidence and key assumptions on both sides,
(q) illustrate one's central concepts with significant examples and show how they apply in real situations, etc.,
(r) empathetically entertain strong objections from points of view other than one's own.[175]

---

[175] Ibid., 4.

The Center's model of critical thinking embraces four perspectives that apply to all critical thinking analysis, what are termed the component domains:

(1) elements of thought
(2) macro-abilities
(3) affective dimensions
(4) intellectual standards.[176]

Paul and Nosich concede that although "all these dimensions are essential, it does not follow that all are directly testable, nor does it follow that any of them is easily testable."[177] The domains:

(1) ***The elements of thought***

> are the basic building blocks of thinking, essential dimensions of reasoning whenever and wherever it occurs. Working together, they shape reasoning and provide a general logic to reason. We can articulate these elements by paying close attention to what is implicit in the attempt on the part of the mind to figure anything out whatsoever. Once we make them clear, it will be obvious that each of them can serve as an important touchstone or point of assessment in critical analysis and in the assessment of thinking.[178]

The fundamental structures of thought serve as the context for applying certain basic thinking skills, characterized as micro-skills, out of which larger-domained critical thinking abilities are built: the ability to identify, clarify and argue for and against alternative formulations of the elements of thought. To the elements of thought students "gather, conceptualize, apply, analyze, synthesize, or evaluate information.[179]

---

[176] The authors' fourth domain rests on the premise that higher order thinking meets certain universal intellectual standards which apply to thinking in every subject. These are summarized in a chart in Chapter Three.

[177] Ibid., 15.

[178] Ibid., 21.

[179] Ibid., 12.

(2) *Macro-abilities*

A number of macro-abilities are invoked when thinking refers to "complex and sometimes ambiguous issues, problems, decisions, theories, states of affairs, social institutions, and human artifacts. These critical thinking macro-abilities include being skillful at:

a. refining generalizations and avoiding over-simplifications,
b. comparing analogous situations: transferring insights into new contexts,
c. developing one's perspective: creating or exploring the implications of beliefs, arguments, or theories,
d. clarifying issues, conclusions, or beliefs,
e. clarifying and analyzing the meanings of words and phrases, [constructing and clarifying interpretations]
f. developing criteria for evaluation: clarifying values and standards,
g. evaluating the credibility of sources of information,
h. questioning deeply: raising and pursuing root or significant questions,
i. analyzing or evaluating arguments, interpretations, beliefs, or theories,
j. generating or assessing solutions,
k. analyzing or evaluating actions or policies,
l. reasoning dialogically: comparing perspectives, interpretations, theories,
m. reasoning dialectically: evaluating perspectives, interpretations,or theories,
n. reading critically: constructing an accurate interpretation of, understanding the elements of thought in, and evaluating, the reasoning of a text,
o. listening critically: constructing an accurate interpretation of, understanding the elements of thought in, and evaluating,the reasoning of an oral communication,
p. writing critically: creating, developing, clarifying and conveying, in written form, the logic of one's thinking,
q. speaking critically: creating, developing, clarifying and conveying, in spoken form, the logic of one's thinking.[180]

---

[180] Ibid., 14.

(3) *Affective dimensions*

Higher order thinking according to this model also involves another crucial perspective which the authors refer to as affective dimensions: certain attitudes, dispositions, passions, and traits of mind. These affective dimensions include:

a. thinking independently,
b. exercising fairmindedness,
c. developing insight into egocentricity and sociocentricity,
d. developing intellectual humility and suspending judgment,
e. developing intellectual courage,
f. developing intellectual good faith and integrity,
g. developing intellectual perseverance,
h. developing confidence in reason,
i. exploring thoughts underlying feelings and feelings underlying thoughts,
j. developing intellectual curiosity.[181]

These elements are not merely important to critical thinking, they are *essential* to the effective use of higher order thinking in real settings. As Boehm points out, the authors appreciate that "for some of these affective dimensions (intellectual perseverance, for example) any testing would have to take place over an appropriately long period of time and thus [they] could not be legitimately assessed at all during the time-frame suitable for a national test." Notwithstanding this caveat, Boehm argues for including the affective component in a national assessment of college student learning [anticipating points to be made later in this chapter by White and the authors from Alverno]: "It seems pretty clear to me that these *can* be assessed very effectively by portfolio, which allows for, even calls for, a variety of assessment materials, including, especially, drafts of essays on a range of topics, which reflect the student's thinking process as well as his disposition toward thinking."[182]

---

[181] Ibid., 15.

[182] Boehm on Paul and Nosich.

Facione reports that the Delphi research consensus also identified "a set of critical thinking dispositions that characterized how a good critical thinker approaches life and living in general and specific problems or questions that might arise, [and in so doing] drew some instructive distinctions."[183]

Table 5 of the Delphi report describes the "Affective Dispositions of Critical Thinking":

1) Inquisitiveness with regard to a wide range of issues,
2) Concern to become and remain generally well-informed,
3) Alertness to opportunities to use critical thinking,
4) Trust in the processes of reasoned inquiry,
5) Self-confidence in one's own ability to reason,
6) Open-mindedness regarding divergent world views,
7) Flexibility in considering alternatives and opinions,
8) Understanding opinions of other people,
9) Fair-mindedness in appraising reasoning, and
10) Honesty in facing one's own biases, prejudices, stereotypes, egocentric or sociocentric tendencies.

An instrument that might capture these affective dimensions thus—at a minimum—will be constructed upon real and practical problems, a premise which points in the direction taken by another advisor.

Practicality as a Litmus Test

While this grand catalog of critical thinking skills and abilities may seem impractically exhaustive, categorically overlapping, and dauntingly complex, a strategy to begin to narrow and refine the target came from another author inside the critical thinking movement, Susan Nummedal, chair of the assessment group of the California State University Critical Thinking Council. She argues that "attention be given to an assessment of the skills and dispositions of practical intelligence and that the assessment measures incorporate the key characteristics of practical thought, including its social nature."[184] She points out [as do others, and as analyzed earlier around Swanson's remark] that Goal 5 strives for several—not necessarily compatible—results. She would probably restate the complete Goal 5 thus:

---

[183] Facione on Paul and Nosich, 4. Quoting from the Executive Summary which describes the experts conclusions, Facione says, "The deepest division is between the nearly two-thirds majority who hold that the term "critical thinking" includes in its meaning a reference to certain affective dispositions and the roughly one-third minority who hold that "critical thinking" refers only to cognitive skills and dispositions, but not to affective dispositions. . . The minority distinguish sharply between what is true of critical *thinking* from what is true of good critical *thinkers*. . . The strict proceduralists do not find it sensible to deny that a person is a critical thinker simply because the person, while skilled in critical thinking, fails to check the credibility of sources, gives up too soon when asked to work a challenging problem, lacks confidence in using reason to approach everyday problems, or ignores painful facts. These experts hold that such a person, because of his critical thinking skills, should be called a critical thinker--but not a good one (in terms of his effective use of those skills)." [5]

> To what extent should the target communicating effectively, problem solving and critical thinking skills (once specified) be further focused or refined to those higher order thinking skills acquired through the college experience that are associated with success in the workforce as well as effective citizenship.

She intends to clarify these disparate elements of the charge by offering a strong suggestion: "I want to argue that we should focus our attention on the latter, namely those skills acquired through learning experiences in college that are relevant to successful functioning in real-life situations."[185]

Nummedal has been actively involved in evaluating how curriculum changes (critical thinking in particular) instituted in American colleges actually work. She states that "when specific courses in critical thinking have been introduced into the curriculum, one of their major goals has been to enable students to think more effectively about everyday issues and concerns, such as practical problem solving and decision making."[186] She cites research undergirding the debate over the value of general reasoning processes vs. domain specific processes, where this element is described variously as transfer, generalization, or application.[187] She suggests putting the horse before the cart by asking a question that amounts to a definitive reframing of the national assessment of college student learning:

> "What is the nature of everyday thinking, practical thought, and practical problem solving?

---

[184] Nummedal, op. cit., 2.

[185] Ibid., 4.

[186] Ibid., 5.

[187] She cites R. Glaser, R. "Education and Thinking: The Role of Knowledge." *American Psychologist, 39(1984):* 93-104.

To begin to answer the question, it is important to distinguish between what Neisser[188] has called academic intelligence (i.e., what it takes to be successful in school settings) and practical intelligence (i.e., what it takes to perform successfully in real world settings)."[189] The fact that Goal 5 makes explicit the element of competitiveness, says Nummedal, suggests that "we need to know what these ingredients of practical intelligence and life success are."[190]

This seemingly revolutionary reshuffling of concepts troubled Facione:

> In summary, as tempting as it might first seem, we do not have to begin the science of critical thinking assessment all over again. . .the historically interesting issues raised in Dr. Nummedal's paper, when compared to the extant research and successful projects, are insufficient to persuade us to abandon decades of work and start over. For scientific progress in critical thinking assessment to continue we must pursue the research agendas now in place. We must build on our successes and respond to the challenges suggested by the objective data on college student abilities now starting to become available. We know what critical thinking is and we have begun to find successful ways to assess it. Let's move ahead.[191]

Moreover, John Chaffee took issue with the significance of a couple of distinctions made by Nummedal, seeming to reinforce Facione's argument and suggesting that emphasizing semantic distinctions should not forestall or scuttle the national assessment of college student learning effort:

> The dichotomy between 'academic' and 'real-world' contexts is problematic. I believe that the author is quite right in insisting that a meaningful assessment of cognitive abilities involves practical contexts and real-world situations. However, *effective* academic instruction does in fact integrate practical, real-world problems, issues, case studies, and learning situations. This integration certainly needs to be expanded and informed by a broad range of learning contexts, but it is not accurate or helpful to

---

[188] U. Neisser, "General, Academic, and Artificial Intelligence." in *The Nature of Intelligence*, ed. L.B. Resnick (New York: Wiley and Sons, 1976).

[189] Nummedal, op. cit. 5.

[190] Ibid., 5.

[191] Facione on Nummedal

> portray the academic arena exclusively as a self-contained, hermetically-sealed, educational biosphere, completely isolated from practical considerations and life experience.
>
> Further, the dichotomy between 'theoretical intelligence' and 'practical intelligence' is also problematic. The fact is that theoretical understanding and practical application function most effectively in dialectical integration, whether the context is academic or non-academic. To paraphrase Immanuel Kant: "Applications without conceptual understanding is blind; conceptual understanding without application is empty." In order to develop complex cognitive and communication abilities in a meaningful and lasting fashion, people need to develop both theoretical, conceptual frameworks *and* the ability to apply these frameworks to practical contexts. This should be a guiding principle for the current NCES project.[192]

Another issue raised by Nummedal's call for practicality and relevance to real life was what such a tilt might exclude. Magda Kohlberg from the U.S. Office of Personnel Management (OPM) reminded the scholars that "real life situations are always attached to contexts." At her agency, "particularly in our job analysis of professional, administrative, higher level positions—the capacity for abstraction and inferential potential in the ambiance of abstraction is extremely important."[193] Added Paul, "on the question of abstractness and putting things in context:

> The kinds of problems that we should assess, that are real world problems, are not ones that are so fixed in a particular context that they are idiosyncratic, but rather those problems that are real world problems that are broad and cross discipline areas. Lots of problems of ecology, for example, involve reflecting in an historical and a political and an economic and a moral sense on the same problem. That is, the problem has many dimensions to it. It is also embedded in a variety of contexts. And this kind of thinking then involves problem solving, involves the use of language in very effective ways, involves critical thinking, and undoubtedly involves background information and other kinds of considerations, which may or may not be put into the prompt itself.[194]

---

[192] Ibid., 2.

[193] Magda Kohlberg in Open Session.

[194] Richard Paul in Open Session.

For Paul, problem solving doesn't constitute a distinct domain because certain intellectual standards underlie all reasoning, and colleges will have incorporated effective critical thinking when students "learn to reason historically, learn to reason economically, learn to reason mathematically, learn to reason scientifically."[195] These universal standards mean that critical thinkers are continually evaluating their own thinking and reasoning for—among other elements—the best problem-solving approaches and the type of reasoning that best suits the problem at hand. He believes strongly that "this is integral to our understanding of what critical thinking is."[196]

Nummedal did make a point that was generally reinforced in workshop discussions—that most critical thinking analyses emphasize the multiple nature of interacting skills:

> no single skill or disposition can be equated with higher order thinking. True, individual skills may be activated in the service of some higher order thinking activity. But when one engages in an activity requiring higher order thinking, a unique subset of skills will be invoked in the service of that activity and may be used and combined in ways unique to that particular activity.[197]

This truth has implications for assessment.[198] She warns that "the mere summation of performance on objectively measurable and highly specific sub-tasks can not be equated with higher order thinking competence," and thus she recommends that "authentic, performance-based tasks should be designed which explicitly incorporate a number of specific skills. Given the complex nature of these tasks, performance on one task may be judged for several purposes corresponding to different skill assessments."[199]

---

[195] Ibid.

[196] Ibid.

[197] Nummedal, op. cit., 12.

[198] Nummedal echoes the term "task" that was used in the GIS in New Jersey. See Edward A. Morante, 1991.

[199] Nummedal, op. cit., 13.

Studies of "on-the-job" practical thinking and problem solving reveal that "both the problems themselves and the practical thinking necessary to deal successfully with them share a number of common features:"[200]

(1) They often tend to be ill-structured or "messy" problems. "Thus, one feature of everyday problem solving is that it requires formulating or redefining the problem itself as well as generating problem solutions. Goodnow[201] has presented convincing evidence documenting the importance of organizing—and, particularly, *reorganizing*—to intelligent behavior in daily life. This would seem to imply that individuals who are successful in solving real-world problems are able to differentiate between well-structured and ill-structured problems, devoting some of their resources to restructuring and reorganizing where appropriate."[202]

(2) "Given the ill-structured nature of everyday problems, it should not be surprising to find that two of the most important features characterizing solutions to these problems [are] efficiency and flexibility."[203] Useful studies of efficiency[204] have been mounted looking at "economy of effort"[205] and "minimizing cognitive load."[206] Flexibility entails the

---

[200] Ibid., 6.

[201] J. J. Goodnow, "Some Lifelong Everyday Forms of Intelligent Behavior: Organizing and Reorganizing," in *Practical Intelligence: Nature and Origins of Competence in the Everyday World*. eds. R.J. Sternberg and R.K. Wagner (Cambridge: Cambridge University Press, 1986): 143-162.

[202] Nummedal, op. cit., 7. According to Ratcliff, this truth about problem structure has largely found its way into the critical thinking orthodoxy. "Assessments of students' critical thinking abilities illustrate this point. One factor differentiating tests of critical thinking is that of problem structure. Problem structure is the extent to which a problem can be described fully and can be answered rightly or wrongly. Complex social, political or economic problems do not have right or wrong answers. Often their very nature is debated. These are ill-structured problem sets. In contrast, problems that can be solved by deductive logic (in the spirit of Sherlock Holmes or Miss Marple) possess a high degree of certainty and correctness. They are well-structured problems." Ratcliff, op. cit., 22.

[203] Nummedal, op. cit., 7.

[204] B. Rogoff and J. Lave (Eds.) *Everyday Cognition: Its Development in Social Context*. Cambridge: Harvard University Press, 1984.

[205] S. Scribner, "Studying working intelligence, " in *Everyday Cognition: Its Development in Social Context*, eds. B. Rogoff and J. Lave (Cambridge: Harvard University Press, 1984).

[206] R.S. Nickerson, D.N. Perkins and E.E. Smith, *The Teaching of Thinking* (Hillsdale, N.J.: Erlbaum, 1985).

ability to shift among solution strategies as the problem space requires. Other elements of this skill include "informal improvisation"[207] and effective "use of the environment (including its social, symbolic, and material resources) in which the problem is situated to effect better problem solutions."[208]

(3) "Finally, there is a very important social component to everyday thinking and problem solving. The model of the individual problem solver sitting down alone to face a problem and come up with a solution independent of input from other people may be very rare indeed—and certainly cannot be assumed to be the first step in problem solving."[209] Nummedal observed that, while Goal 5 refers explicitly to American competitiveness in the global economy, other values may be implicit—in the task, if not the stated goal. "We need to incorporate into the assessment skills [that are] associated with cooperation. *America 2000* speaks of a competitive workforce. Nowhere does it speak about the importance of a *cooperative* one. I believe we should focus at least as much on measuring the success of our institutions of higher learning in promoting cooperation as in promoting competition."[210]

How can these target skills be incorporated into a national assessment of college student learning? Nummedal does not see much hope in extant efforts presently in the colleges. "For the most part, I have the same reservations about the [current critical thinking] tests [such as Cornell and Watson-Glaser] as I have about existing critical thinking courses as guides for selecting the skills and dispositions to be included in this assessment. I think in both cases, the skills and dispositions are derived from conceptions of critical thinking that are both narrow and discipline specific."[211] She advises that "we need to look to the world of practical thinking and identify those skills and dispositions associated with successful performance. I believe we should build on the experiences of those who already have been about the business of creating authentic, performance-based measures,[212] examining these measures for higher order thinking skills and dispositions that appear to be associated with performance in real world situations."[213]

---

[207] D.A. Schon, *The Reflective Practitioner* (New York: Basic, 1983).

[208] Nummedal, op. cit., 7.

[209] Ibid., 8.

[210] Ibid., 16.

[211] Ibid., 11.

[212] For example, D.A. Archbald and F.M. Newmann, *Beyond Standardized Testing: Assessing Authentic Academic Achievement in the Secondary School* (Reston, VA: National Association of Secondary School Principals, 1988).

Assessment in the Workplace

Daniel Resnick and Natalie Peterson echoed many of Nummedal's concerns in reminding their colleagues that it is the American business community whose stake in Goal 5 is directly addressed by the reference to a competitive workforce. Their "leadership has been challenged in world markets by nations with better educated workforces that are exploiting economic opportunities more effectively."[214] If the call for performance in real world situations is heeded, they suggest, the workplace is the most pragmatic context toward which to target the search for skills and abilities.

Resnick and Peterson emphasize that whatever results are achieved in colleges, they must be manifest when graduates begin to work. That is, the skills and abilities must *transfer*.[215] They contend that "no set of indicators for goal five can afford to omit the perceptions that employers have of college students as they enter the workforce. An annual employer survey would raise the salience of employers' expectations for college students at the same time that it would measure progress toward greater work readiness."[216] They cite a 1991 Louis Harris and Associates study as indicative of the proper approach that embraces many diverse viewpoints:

> At the core of the employer, educator, parent, and former student protocols was a set of items requiring the participant to rate recent high school graduates on a common set of 15 attributes related to a young person's ability to perform well in higher education or on the job. . . . Four of these attributes concerned basic skills in reading, writing, and mathematics. Two attributes--the ability to solve complex problems and the

---

[213] Nummedal, op. cit., 12.

[214] Resnick and Peterson, op. cit., 2.

[215] Although Resnick and Peterson report that "there is some disagreement in the scholarly literature about how much education the American workplaces of the future will require of their workers. John Bishop, [writing] "A Worsening Shortage of College Graduates," in *Educational Evaluation and Policy Analysis* (forthcoming), reviews these arguments." [2]

[216] Ibid., 16.

ability to read and understand written and verbal directions--were higher order in nature. The remaining attributes described personal qualities such as having a good attitude and knowing how to dress and behave.[217]

Following this lead, Resnick and Peterson catalog the skills needed by the worker of the future that were outlined in the *SCANS* report for *America 2000*,[218] which fall into five domains:

> (1) The first involves the uses of resources. In tomorrow's workplace, employees will be repeatedly called upon to schedule time, budget funds, assign space, and arrange staff.
>
> (2) The second. . .concerns interpersonal skills. Future workers must be adept at working in a team, teaching others new skills, serving clients and customers, exercising leadership, negotiating, and dealing with diversity.
>
> (3) The third area of competency focuses on the ability to use and acquire information. Employees will be expected to acquire, evaluate, organize, maintain, interpret, and evaluate information. They should also be prepared to use computers to process this information.
>
> (4) The fourth domain is concerned with systems. Employees must have an understanding of how social, organizational, and technological systems work and be able to operate effectively with them. Based upon this knowledge, they should be able to monitor and correct performance and improve or even design systems.
>
> (5) The final area of competency involves technology. In the workplace of the future employees [are] expected to have a familiarity with a variety of technologies so that they will be able to select, apply, and maintain them.[219]

---

[217] Ibid., 17.

[218] U.S. Department of Labor, The Secretary's Commission on Achieving Necessary Skills, *What Work Requires of Schools: A SCANS Report for America 2000* (Washington, D.C.), 1991.

[219] Resnick and Peterson, op. cit., 6.

Each of these domains, continues their summary of *SCANS*, "is based upon a foundation of basic skills, thinking skills, and personal qualities:

(1) Basic skills include, at some level, reading, writing, mathematics, listening, and speaking.

(2) Thinking skills, sometimes referred to as higher order skills, are those associated with thinking creatively, making decisions, and solving problems. Related skills include knowing how to learn, reason, and organize as well as to process symbols, pictures, graphs, objects and other information.

(3) Personal qualities, described by the report, include responsibility, self-esteem, sociability, self management, and integrity.[220]

The Current State of Workplace Assessment

Thus do Resnick and Peterson show that Nummedal's call for assessing practical skills is consistent with the domains outlined in the *SCANS* report. But given the criticism by Nummedal and others about the extant college assessments of such skills, a logical inquiry would be to see how the American business community presently evaluates its workforce. As the Co-director of the National Center on the Educational Quality of the Workforce, Peter Cappelli is in a position to report on precisely that. His paper intends to summarize "what can be learned from the American industry experience in analyzing jobs and testing employees that might help advance the goal of assessing and improving college performance."[221]

Cappelli sees "two areas where industry assessments are most applicable to the National Goal of improving and assessing the performance of students in college."[222] The second is with actual employee assessments. The first is with efforts to identify the knowledges, skills, and abilities (KSAs) that are required for jobs, an effort typically referred to as job analysis (a phrase that usually refers to systematic efforts to collect information about the work requirements associated with particular jobs). "The results of job analyses are helpful in that they suggest what employees need to bring to a job in order to be successful. They also suggest the areas where colleges should be preparing students and, in turn, some of the learning that might be tracked in an evaluation scheme."[223]

---

[220] Ibid., 6.

[221] Cappelli, op. cit., 1.

[222] Ibid., 4.

Thus are the KSAs relevant to that aspect of Goal 5 which refers to the work force (competitiveness in the global economy). While several other advisors speculate (some explicitly) on the issue of transfer—that is, skills and abilities that are part of college education which are *believed* to be relevant to required workplace skills—Cappelli's material derives directly from this arena. The framework he concentrates on describes tasks "from the perspective of the worker and describes what is needed from workers in order to perform a given job. The latter is clearly the more useful for the purposes at hand as it describes what jobs demand from workers."[224] Cappelli offers a representative sampling of seven from among the dozens of prominent job analyses currently in use in the workplace:

(1) The *Hay Associates Profile System* "job analysis focuses on three areas:

  (a) "Know-how" concerns the techniques and procedures required by jobs. Examples of know-how would be professional skills, such as accounting or engineering, and general management skills such as designing plans. More specialized and technical skills and greater breadth required across skills is associated with more difficult jobs.

  (b) "Problem solving" refers to the thinking demands made by jobs. Routine, repetitive tasks fall at the lower end of this scale while those defined only abstractly, requiring adaptive abilities, fall at the upper end.

  (c) "Accountability" refers to the freedom jobs give employees to act. Jobs that offer employees little guidance and that also are associated with large impacts on the organization score high on this scale."[225]

---

[223] Ibid., 4.

[224] Ibid., 5.

[225] Ibid., 5.

(2) "*The Position Analysis Questionnaire (PAQ)* has been the most thoroughly researched and academically prominent of the job analysis methods,"[226] consisting "of 187 items in the questionnaire which can be divided into six general categories: Information (where and how one gets information needed for the job), mental processes (reasoning, decision making, etc.), work output (physical activities, tools, etc.), relationships with others (measures of complexity), job context (social and physical context of work), and a catch-all "other" category. While the PAQ's focus on work behaviors, as opposed to tasks, it has sometimes been criticized in the context of differentiating jobs, it is an advantage here in helping to identify what workers need to know."[227]

(3) *The Management Position Description Questionnaire*,[228] specifies knowledge, skills and abilities that can be categorized as follows: "leadership skills (motivation, coaching), administrative skills (planning, allocating), interpersonal skills (conflict management, group process skills), communications, decision making (information management, analytic ability), and professional knowledge (company-specific practices, technical skills such as accounting)."[229]

(4) *The Threshold Traits Analysis System* focuses explicitly on individual job holders, rather than on the jobs themselves, and examines the traits that they possess.[230] "Those traits can be broken down into ability factors, which are subdivided into aptitudes for acquiring knowledge or skill and proficiencies for skills already possessed; and attitudinal factors, which affect the willingness to perform at given

---

[226] Ibid., 6. Cappelli also refers to a Professional and Managerial Position Questionnaire (PMPQ), which he states "is very similar." See Ernest J. McCormick and P. Richard Jeanneret, "Position Analysis Questionnaire," in *The Job Analysis Handbook for Business*, Industry, and Government, ed. Sidney Gael (New York: John Wiley and Sons, 1988): 825-842.

[227] Ibid., 6.

[228] Developed, says Cappelli, "by Control Data Business Advisors for use with their own managerial employees but has become popular in many white collar organizations, in part because its focus on managerial jobs made it appear more applicable to them." [p. 6] See Ronald C. Page, "Management Position Description Questionnaire," in *The Job Analysis Handbook for Business*, Industry, and Government, ed. Sidney Gael (New York: John Wiley and Sons, 1988): 860-879.

[229] Cappelli, op. cit., 6.

[230] Felix Lopez, "Threshold Traits Analysis System," in *The Job Analysis Handbook for Business*, Industry, and Government, ed. Sidney Gael (New York: John Wiley and Sons, 1988): 890-901.

levels. The specific traits. . .are categorized as follows: Physical traits such as strength, mental traits such as problem-solving and memory, learned knowledge and skills such as communication, motivation and adaptability, and social traits such as influence and cooperation."[231]

(5) *Ability Requirement Scales.* "These scales attempt to identify generic abilities and are based on 50 item categories. . .Perhaps more than the other job analysis systems described here, the Ability Requirements Scales focus on physical and perceptual factors. Among the nonphysical categories, communication skills, reasoning, and problem-solving feature heavily."[232]

(6) *The Functional Job Analysis* "is a straightforward description of worker characteristics required for the jobs described in the *Dictionary of Occupational Titles,* and has seven categories: Data functions (complexity in the use of information), people functions (level of interpersonal skills demanded), functions using things (physical requirements, typically with machines), worker instructions (level of responsibility), reasoning development (from common sense to abstract undertakings), mathematical development (math skills), and writing functions."[233]

(7) *SCANS—The Secretary's Commission on Achieving Necessary Skills* "identified five sets of general competencies required by entry level jobs: those associated with resources (organizing, planning, allocating), interpersonal skills, using and acquiring information, understanding systems, and working with technology. Underlying those competencies were three sets of what the Commission called 'foundations.'"[234]

---

[231] Cappelli, op. cit., 7.

[232] Ibid., 7. See Edwin Fleishman and Michael D. Mumford, "Ability Requirement Scales," in *The Job Analysis Handbook for Business*, Industry, and Government, ed. Sidney Gael (New York: John Wiley and Sons, 1988): 917-935.

[233] Ibid., 7. See Fine, op. cit.

[234] Ibid., 8. These foundation skills were listed earlier by Resnick and Peterson. Cappelli continues: "The public policy action with the most widespread impact in this area may be the efforts by the U. S. Department of Labor to identify the requirements of jobs that come through its Employment Services. The criterion used to determine what is demanded from workers includes the General Educational Development (GED) levels in reasoning, math, and language; specific vocational preparations; aptitudes and temperaments; and physical demands." See Robert C. Droege "Department of Labor Job Analysis Methodology," in *The Job Analysis Handbook for Business*, Industry, and Government, ed. Sidney Gael (New York: John Wiley and Sons, 1988): 993-1018.

Cappelli's survey of these extant systems shows that "several requirements cut across virtually every system of job analysis. They include the following sets of KSAs:

(1) Interpersonal skills,
(2) Communications, both oral and written,
(3) Critical thinking broadly defined (problem solving, etc.),
(4) Motivation and other personal attitudinal characteristics,
(5) Working with data and information, and
(6) Math skills.[235]

Do these KSAs figure prominently and explicitly in most of the extant, industry job analysis schemes? Not really, concludes Cappelli, but "most of the knowledge, skills and abilities in the above list are taught in college, albeit some indirectly."[236] He doesn't believe that implementing these KSAs in college instruction requires a revolution so much as a re-orientation from current trends where applications and practical problems are insufficiently stressed.

Several reviewers and workshop participants believed this interface between industry and the colleges was a fruitful area to explore. Boehm pointed out that

> The *SCANS* Report is only the most recent voice in a growing chorus. *Fortune* magazine generally, and the Spring 1990 "Saving Our Schools" issue in particular;[237] the U. S. Department of Labor's *Workplace Basics: The Skills Employers Want*;[238] Motorola's *The Crisis in American Education*;[239] Rockwell's

---

[235] Ibid., 8.

[236] Ibid., 9.

[237] "Perhaps the single best introduction to the growing role of 'the workplace' in education is this Spring 1990 special issue of Fortune. Back issues are still available: Time & Life Building, Rockefeller Center, New York, NY 10020-1393."

[238] Carnevale, Anthony, P., Leila J. Gainer, and Ann S. Meltzer, *Workplace Basics: The Skills Employers Want*, published jointly by The American Society for Training and Development (ASTD) and the U. S. Department of Labor, 1988. Copies are avail- able from ASTD, 1630 Duke Street, Box 1443, Alexandria, VA 22313.

[239] Available from Edward Bates, Director of Education--External Systems, Motorola Inc., 1303 E. Algonquin Road, Schaumberg, IL 60196.

Emphasize Education. It's Our Future;[240] *Workforce 2000: Work and Workers for the Twenty-first Century*;[241] and a variety of other publications have all made quite clear just what 'the workplace' wants from education. The extraordinary and growing interest which college and university faculty in general, and vocational and technical college faculty in particular, have shown in critical thinking indicates that at least they are paying close attention.[242]

Elinor Greenberg noted that while Cappelli's analysis doesn't really produce a certified list of skills, it does suggest "that *one strategy*, among a number of strategies, could be to present to all U.S. colleges (3500+) and, perhaps, to public and private vocational schools (9000+) as well, an opportunity to *choose* a particular approach to assessment that would give us *clusters of institutions*, or *states*, along with like-minded *businesses* with which to partner."[243] Her "Coordinated Multi-Option National Assessment and Partnership System" would encourage the development of related systems which were nonetheless targeted toward a particular interest. With respect to the "Industry-based" version, she suggests a number of elements:

1. Collect *instruments widely and effectively used by industry*, compile a database and directory of such instruments, select a few model sub-skill schemes, and disseminate this information to colleges for adoption and use in courses and in entry, mid-point and exit student assessment processes.

2. Use the above in *faculty development* activities to orient faculty to industry's procedures and cultures and to encourage the use of sub-skills as course objectives and assessment criteria.

3. Develop a model "*college transcript-as-career passport*" to use as a tool for students, colleges and employers to document lifetime learning in schools, on-the-job and in other settings. This *Career/Education Passport* can be input into a computerized

---

[240] Available from Rockwell International, P.O. Box 905, El Segundo, CA 90245-0905.

[241] Johnston, William B. and Arnold H. Packer, *Workforce 2000 Work and Workers for the Twenty-first Century* (The Hudson Institute, 1987).

[242] Boehm on Paul and Nosich.

[243] Greenberg on Cappelli.

database and available in hard copy form. It should contain a wide variety of information on the learner: bio-data, college course credits, grades; degrees; on-the-job training records; competency outcome statements; test scores, etc. The student/worker would "own" the Career/Education Passport. It would be transportable from school to school, job to job and career to career, throughout one's lifetime.

4. The *Career/Education Passport* could also be contained in a chip that is part of a *"smart card,"* to be used as a transfer, registration, and tuition payment device to encourage and simplify recurrent lifelong learning and enrollment. Lifelong learning, thereby, translates into lifelong training and lifelong education through a commonly used, efficient, and well understood technological tool. This "smart card" would help to create an "American Lifelong Learning System" and would simplify the now complex and barrier-filled recurrent, lifelong entry and re-entry to formal learning. This kind of a tool could help create the kind of "seamless," integrated learning system now being widely discussed, but not yet a reality.

5. *Career/Education Assessment Partnerships* between schools and employers could be created, based on shared projects built around items 1-4 above. These "assessment partnerships" would provide a basis for collaborative and mutually supportive arrangements between the academic, business and labor sectors.[244]

Greenberg asserts that the "technology to develop the above items is now available, but not being widely used in the U.S. Issues of cost, attitude and feasibility should be investigated further."[245] Thus is there a fairly rich but unorganized network of ideas and systems extant for assessing the actual needs and status of the workplace. As to the current state of assessment in American colleges, another advisor spoke directly to the recent experience.

---

[244] Ibid.

[245] Ibid.

Assessment in the Colleges

Nummedal and others found little good to say about the current state of critical thinking assessment in American colleges. Charles Lenth's ecumenical survey of that scene, however, is informed by his experience as executive director of the State Higher Education Executive Officers, based in Denver. Lenth's survey of actual state experience with assessment confirms that a majority of states are now (or have been) experimenting with some sort of assessment. However, no coherent picture or consensus emerges about the target skills and abilities, in large part because the motives for the assessments—and the uses to which they been put—vary widely,[246] and rarely have they addressed in any direct way the critical thinking skills specified by the Goals Panel as Objective 5 abilities.

Some examples: Five states have developed and use a common instrument for basic skills and placement testing. The Georgia Regents' Test was developed in the late 1960s following rather extensive research and pilot testing of commercial instruments and system-wide consultation. After considering more extensive testing, the instrument was limited to reading comprehension and writing. The College Placement Examination (CPE) tests basic skills and competencies in the areas of reading, mathematics and English. Arkansas and Vermont also test for placement purposes.

In Texas, since the fall of 1989, entering students have been required to take the statewide examination in reading, writing and mathematics prior to completing nine credit hours of college-level coursework. Ron Swanson is associate director of the Texas Academic Skills Program, and points out that such a structure puts the focus squarely on "precollegiate education and its quality. . .and has two primary goals: [first, to] diagnose academic deficiencies early on and get students the help they need, and [second, to] feed assessment results, grades, etc. back to the high schools from which the students graduated in the hope that reforms will take place in K-12 education."[247] He reports that, despite its recency, "changes in K-12 are already underway." He was one of the more persistent at the workshop in trying to step outside of the framework of Goal 5 to examine "the issue of where, in the educational process, are higher order thinking skills learned or developed?

---

[246] Ibid., 9-12.

[247] Swanson on Lenth.

Knowing that will have great impact on what, if anything, we do with the results of a national assessment."[248]

Lenth looks fairly closely at the Washington state experience, which has been more ambitious than most, and which might help to illuminate "the process and usefulness of multi-institution postsecondary testing."[249] Task forces were established, which then identified three national tests for measuring college-level critical thinking, computational and communication skills (ACT-COMP, ACT-CAAP and the ETS Academic Profile), and set up procedures to pilot test and evaluate these extant measures. Ultimately, reports Lenth, several years worth of pilot testing and evaluation produced the conclusion, shared by faculty, that "None of the three tests was judged a valid and useful measure of the three skill areas."[250]

Given the patchwork of state systems throughout the country, Lenth concludes that observers might view the glass as either half full or half empty. As Lenth concedes, and his reviewer Larson emphasizes:

> his paper, though affirmative in what it *says,* is highly cautionary in what it *expresses;* he could almost as easily have reached the conclusion opposite to the one he presents: that the effort he discusses is doomed to failure. I was struck in particular by Lenth's report that, although interest in assessment has picked up over the last few years, many states direct *institution-based* assessment instead of statewide procedures, and begin their assessment with basic skills and with the qualifications students require in order to pass through gateways, instead of focusing on ultimate outcomes.
>
> I was further struck by the twin recognitions that outcomes-based assessment of college graduates is untried and that it will be exceedingly difficult to conduct. I noted that Dr. Lenth made no effort whatever to suggest a specific educational context for assessment or even the embryonic outlines of a procedure for assessing; Dr. Lenth, as noted, is describing a policy and administrative problem, rather than an

---

[248] Ibid.

[249] Lenth, op. cit., 14.

[250] Ibid., 18.

> approach to assessment. And his reviews of current history demonstrate the urgency of that problem.[251]

Larson interprets the national experience as summarized by Lenth to suggest that "working with the states seems to offer only a marginal possibility of success." He points out that Lenth provides a real warning, leaving for later the question of "*how* to work with the states, or how to organize people for discussion of ways to collaborate." This major caveat, for Larson, promises that the ways to deal with such problems, if and when found, "will be expensive."[252]

Basic Skills First, then General Intellectual Skills

Lenth, however, chooses to take the more constructive view, and proposes that the national assessment of college student learning be conceived in such a way as to avoid duplication and hence any institutional rivalries with the various states. His survey supports his conclusion that "ongoing state and institution assessment activities focus primarily on the teaching and learning environment,"[253] which makes them inherently unsatisfactory as measures of general outcomes. Similarly, many of the current systems target basic skills and establish "expected competency" levels because such benchmarks are necessary to establish any sort of coherent context for assessment. Since these needs are so integral to a functioning system, it is probably best that they be left to the individual states and institutions to devise and implement, Lenth believes. In fact, the national assessment of college student learning can function best, he believes, by "developing reliable measures of general intellectual skills;" in particular, an assessment of "*higher order intellectual skills* . . . that help to define and measure the highest levels of academic performance."[254]

General Intellectual Skills (GIS) were the focus—in fact the name—of the test developed in New Jersey. Another of the NCES advisors was Ed Morante, who for a decade through

---

[251] Larson on Lenth.

[252] Ibid.

[253] Lenth, op. cit., summary page.

[254] Ibid., summary page. While many of his colleagues agree that the standards of NACSL should be high, many also believe that general outcomes indicators cannot reveal the highest order of thinking, and that NACSL must be discipline-specific. This issue is addressed in Chapter Three on Standards.

1991 served as Director of two successive testing programs for the New Jersey Department of Higher Education, and who analyzed that experience in view of a national assessment of college student learning. "For more than a dozen years [beginning in 1977], students enrolled in public colleges and universities in New Jersey have been assessed using a common statewide instrument, the New Jersey College Basic Skills Placement Test, [which] assesses the basic skills (reading, writing, and mathematics) of entering college students."[255]

The Basic Skills Council created to help develop the test reached a consensus on a definition:

> By "basic skills" the Council means the tools of intellectual discourse used in common by participating members of all academic communities. These tools are the language of words and the language of mathematics. Students need these tools to extract information, to exercise and develop the critical faculties of the mind, and to express thoughts clearly and coherently.
>
> Without them, learning is impaired, communication is imprecise, understanding is impossible. A test of "basic skills," therefore, is a test to determine whether an individual has developed the practical working skills of verbal and mathematical literacy needed to take advantage of the learning opportunities that colleges provide.[256]

Further, the Council anticipated some objections:

> To define "basic skills" in this way is not to deny the validity of other modes of communication—within the artistic realm of discourse, for instance, the languages of music, motion, image, color, light, and texture express a universe of perceptions, feelings, and emotions which cannot be expressed adequately by words and numbers and logic alone. Nor is the Council's definition of the "basic skills" inimical to the value of diversity. We are, to the contrary, exceedingly sensitive to the differences between colleges: differences in their students; differences in their curricula and

---

[255] Morante, op. cit., 1.

[256] Ibid., 1. See New Jersey Basic Skills Council, *Report to the Board of Higher Education on the Results of the New Jersey College Basic Skills Placement Testing, Fall 1990 Entering Freshmen* (Trenton, NJ: Department of Higher Education) 1991).

> pedagogical philosophies; differences in their missions. But in one respect all colleges are identical: their ultimate purpose is to foster learning. The Council asserts unequivocally that the "basic skills" of reading, writing, and mathematics are a prerequisite to learning at the college level. If the possession of these skills is "standardization," we believe that standardization in this sense is good.[257]

These criteria soon resulted in the establishment of the New Jersey College Basic Skills Placement Test (NJCBSPT), which currently[258] has five components: Reading Comprehension; Essay; Sentence Sense; Computation; and Elementary Algebra.

Lenth concluded that states are best equipped to establish such basic skills assessments, and then to use them for several important functions in their infrastructure and state educational system. Such a foundation, he suggested would probably have to precede any assessment focused on higher-order thinking, which he believes should be outcomes-oriented. The New Jersey experience described by Morante would seem to confirm this relationship, because once the NJBSPT was firmly established, a new statewide assessment effort called the College Outcomes Evaluation Program (COEP) was inaugurated. COEP featured the development of a "sophomore test" labelled a GIS (General Intellectual Skills) Assessment.

These general intellectual skills were the equivalent of the traditional skills of critical thinking, problem solving, quantitative reasoning, and communications (both oral and writing). Three definitions were established, and are elaborated in detail in the Appendix to Morante's paper:[259]

(1) *Accumulate and Examine Information (Gathering Information)*—including the skills necessary to: determine the kinds of information needed for a given task; construct and implement systematic search procedures using both traditional and computerized methods; discard or retain information based on an initial screening for relevance and credibility; and develop abstract concepts appropriate to the task at hand for initially ordering the information which is retained.

---

[257] Ibid., 1.

[258] "Logical Relationships" was dropped as an element after three years when analysis revealed that it was too closely related to the reading and writing components of the test, reports Morante.

[259] Ibid., 27-30.

(2) *Reconfigure, Think About, and Draw Conclusions from Information (Analyzing Information)*—including the skills necessary to: evaluate the interpretations presented by others in terms of their assumptions, logical inferences, and empirical evidence; reconfigure information in ways that suggest a range of alternative interpretations and evaluate their relative merits; construct hypotheses that logically extend thought from areas in which information is already available into areas where it is not; specify the additional information which might confirm or disconfirm those hypotheses; and draw conclusions based on all of the above.

(3) *Present Information (Presenting Information)*—including the skills necessary to express one's own ideas in written, oral, and graphic forms which will be intelligible and persuasive to a variety of audiences.

(4) As a result of test development, a fourth area was added for scoring purposes: *Quantitative Analysis* which "replicates analyzing information but concentrates on problems requiring quantitative reasoning and calculations."[260]

The test has been developed, piloted, and refined in subsequent years, and now consists of 14 separate tasks, a randomly chosen seven of which will appear on a given administration of the test. Given to all students at the 31 public two- and four-year colleges in New Jersey in both 1990 and 1991, "the emphasis is on assessing the underlying general intellectual skills needed by all students regardless of major or institution."[261]

Possible Models at Individual Schools

The most substantive pair of papers with respect to a viable, functioning postsecondary assessment system came from Marcia Mentkowski and Georgine Loacker at Alverno College in Milwaukee, Wisconsin. Unique among American colleges, Alverno (for women) provides an 18-year institutional experience with what the authors describe as a "performance-based, outcome-oriented approach to liberal arts education. To earn a degree, a student demonstrates eight broad abilities: communication, analysis, problem solving, valuing in decision-making, effective interaction, responsibility toward the global environment, effective citizenship, and aesthetic response at increasingly complex levels in a wide variety of settings and contexts."[262]

---

[260] Ibid., 14.

[261] Ibid., 17.

Beyond this breakdown, students develop an understanding of three broad domains of knowledge: natural sciences and mathematics; behavioral and social sciences; and arts and humanities. The major skills, demonstrable in the knowledge domains, are continuously being evaluated with respect to six major performance characteristics: integration; independence; creativity; awareness; commitment; and habituality.

Mentkowski is a psychology professor and the Director of Research and Evaluation at Alverno, and to some extent her paper assumes the reader's awareness of her colleague's description of the Alverno system. Alverno's program is unique in a number of ways, one being that assessment is an integral part of the process throughout a student's career, a dialogue that continues after graduation and throughout the graduate's subsequent work life. The richness of data and experience this structure provides is indisputably incomparable, at least in the American experience.

On the other hand, the extent to which the lessons learned and the principles that have been developed there can translate to the present exercise was to become a strong focus of discussion. As with several other advisors, Mentkowski more or less categorically rejects as too narrow the concept of a single instrument assessment as charged in Goal 5. Specifically, "We assume that readers accept the idea of a national assessment system rather than a single national test, and that they are persons in the process of considering what it means to assess the abilities of students during and after college."[263] The credo behind the target abilities at Alverno specifies abilities that: involve the whole person, are teachable, can be assessed, transfer across settings, and can be continually re-evaluated and defined.

While these aspects of skills and abilities probably do reflect a consensus emerging from the NCES workshop process, Alverno's net goes much wider. The performance characteristics mentioned above are but only the first level of more sophisticated outcomes produced by the Alverno assessment system. More controversial with respect to their relevance to a doable national assessment of college student learning may be such "special abilities" as learning how to learn, and understanding oneself and developing skills of self-assessment.

---

[262] Loacker, op. cit., 1. The paper provides a definitive overview of the structure of the curriculum and the process by which it is conducted. Appendix A further describes a series of six levels for each of these abilities, moving to progressively more complex levels. Appendix B lists the "Advanced Outcomes in the Major Areas at Alverno College." For each of the three majors that can be elected, biology, english, and business & management, a series of abilities are identified that characterize the student who has participated in that major.

[263] Mentkowski, op. cit., 2.

Nonetheless, many of the principles and concepts that she explains as they are exemplified by the Alverno system relate to the primary issues around a national assessment of college student learning. Clearly, some major translation would be in order to implement them in the narrower, less ambitious, and more limited framework that may constrain practical decision-makers. An important point to stress in defense of the relevance of the Alverno model, however, is how integrated it is with the world of work.

A summary of some of these concepts as they relate to the skills and abilities question: a way of discerning important abilities is to discover those which connect education and work: outcomes are abilities that transfer to performance at work; "learning to learn" skills that facilitate developing other abilities; using abilities to create a theory of action; using abilities to achieve and experience a sense of competence; defining abilities with multiple components that are integrated, developmental (teachable) and transferable across contexts and settings; such multiple components include motives or dispositions, self-perceptions and attitudes, skills, behaviors and knowledge; components of complex abilities may include cognitive, affective, behavioral, motivational and perceptual components.[264]

Another program already in place is The Creative and Critical Thinking Program at LaGuardia Community College at The City College of New York. John Chaffee, Director of Creative and Critical Thinking Studies, reports the program "is based on the assumption that thinking is a process that can be understood and improved through proper study and practice."[265] The program has earned grants from The National Endowment on the Humanities (NEH), and is "designed to integrate critical thinking abilities across the college's curriculum. . .Students involved in the program have consistently demonstrated improved thinking abilities and accelerated development of language skills."[266]

---

[264] Ibid., pp. 8 - 17.

[265] Chaffee on Nummedal.

[266] Ibid., 1.

Chaffee cites an evaluation by ETS, as detailed in "The Final Report to NEH"[267] which concludes that the program appears to have succeeded in meeting its primary objectives:

(1) Reasoning and Problem Solving—Utilizing a variety of evaluation strategies, the major evaluator of the project, Dr. Garlie Forehand, Director of Research, Program-Planning and Development at the Educational Testing Service concluded that the program fosters the development of students' thinking abilities at both general and specific levels. At the general level, teachers perceive more respect for the thinking process, more tendency to bring a "habit of thinking" to their classes. At the specific level, teachers reported instances of transfer of such skills as breaking problems into parts, classifying, organization of thought, asking questions, separating facts from opinions, and assessing points of view.

(2) Literacy—Since language and thinking are such closely related, reciprocal and interactive processes, the LaGuardia program is designed to improve students' thinking abilities while simultaneously enhancing their language skills. The cumulative results of the program have revealed that students enrolled in *Critical Thinking Skills* pairs have consistently demonstrated accelerated development of language skills as measured by standard language examinations. In addition to improvements in students' grammatical and structural language skills, faculty also report that students are learning to use language with a depth, insight and sophistication unusual for students at this level, as they seek to utilize and express their evolving higher-order thinking abilities.

(3) Critical Attitudes—One of the guiding principles of the Creative and Critical Thinking program is the belief that learning should take place in experiential contexts, serving to stimulate qualities such as self-awareness, initiative and maturity. . . .Becoming a critical thinker does not simply involve developing discrete intellectual abilities: it involves developing insight, reflective judgement, informed beliefs and a willingness to carefully explore diverse perspectives with incisive questions. As students develop their critical thinking abilities, they also grow as individuals, developing the qualities of openmindedness, responsibility, initiative, and a sense that they can control the direction of their lives through the choices that they make.[268]

---

[267] Available by writing to Chaffee at 31-10 Thomson Avenue, Long Island City, NY 11101.

[268] Amendment to Chaffee on Dunbar.

Frederiksen's view of an improved test structure is not far from this model, and also seems consistent with the GIS [New Jersey] task formulation. Citing Banta and others' support of a new examination system that would emphasize alternatives to the traditional multiple-choice format, he prefers "to see developed tests in the form of realistic simulations of real-life problem situations in the various disciplines. The responses might be a statement of what the examinee would do or say rather than choosing options, as in a multiple-choice test,"[269] and he cites as an example a set of "Tests of Scientific Thinking" intended for graduate psychology students:

> One of the tests is called "Formulating Hypotheses" (FH). Each FH problem requires the examinees to (1) read a brief description of an experiment; (2) study a graph or table showing the results of the experiment; (3) read a statement of the major finding; and (4) write hypotheses (possible explanations) that might account for the finding. The problem has no single right answer, but there are many possible answers that vary widely in quality. The scoring system involves (1) making a classification of the ideas written by the students who took the test, thus forming a set of mutually exclusive categories, and (2) having the categories valued by expert judges. Scoring then involves assigning each response to one of the categories and letting the computer do the rest.[270]

Frederiksen is among those who favor the discipline structure already being taught in colleges over inaugurating a new effort focused on Objective 5 abilities. "The test format should not be of the multiple choice variety; instead, it would be best to create tests that are realistic simulations of situations that elicit relevant behaviors. Problems so posed could increase in difficulty as the college *years* go by, with increasing need for "higher-order thinking skills."[271]

---

[269] Frederiksen on Banta.

[270] Ibid. See N. Frederiksen and W.B. Ward, "Measures for the study of creativity in scientific problem solving," in *Applied Psychological Measurement* (1978): 1-24; and

W.C. Ward, N. Frederiksen, and S.B. Carlson, "Construct validity of free-response and machine scorable forms of a test," in *Journal of Education Measurement*, 17(1980): 11-29.

[271] Frederiksen on Dunbar. See S.F. Chapman, "The higher-order cognitive skills: What they are and how they might be transmitted," In eds. T. C. Sticht, B. A. McDonald, and E.M.J. Beeler, *The intergenerational transfer of cognitive skills*, (Norwood, NJ: Ablex, 1990).

Literacy and Writing Assessments

Literacy is the primary focus of Richard Venezky's contribution, considered as an example of how to approach assessing higher order thinking and communication skills. He concedes that the tasks in this domain don't embrace the totality of skills and abilities specified in Objective 5, but "with some squeezing, these tasks might be aligned with the Delphi classification of core critical thinking skills."[272] Thus, as both an example of the type of problems one may encounter in trying to design a national assessment, and a head start on the literacy component of the national assessment of college student learning, his survey goes into literacy assessment in some depth.

A first hurdle is the definition of literacy, which has often been a functional one, but which can also be cognitive, says Venezky, by attempting "to define literacy in terms of levels of expertise for human performance."[273] The former has been far the more common, but Venezky recommends the latter approach for the national assessment of college student learning, and provides some background on earlier research into the cognitive assessment of literacy.[274] Next, what are the actual skills to assess for college-level literacy? History, culture and education have brought America to the point where basic literacy now involves reading, writing, and basic levels of arithmetic ability. "Even within these three areas, however, issues of assessment coverage exist. Should a college-level literacy assessment include vocabulary assessment? Should it include separate assessments of charts, graphs, tables, and the like? And should it include creative writing as well as expository writing?"[275]

---

[272] Ibid., 1.

[273] Ibid. 4.

[274] See E.L. Thorndike, E.L., The understanding of sentences: A study of errors in reading. *Elementary School Journal,* 17(1917): 98-114;

K. Neijs, *Literacy primers: Construction, evaluation and use.* (Paris: UNESCO, 1961);

R. Glaser and D.J. Klaus, Proficiency measurement: Assessing human performance, In ed. R. M. Gagne, *Psychological principles in system development.* (NY: Holt, Rinehart & Winston, 1962);

P.D. Pearson, ed., *Handbook of reading research.* (NY: Longman, 1984);

E.L. Baker, M. Freeman, and S. Clayton, "Cognitive assessment of history for large scale testing," In eds. M.C. Wittrock & E.L. Baker, *Testing and cognition,* (Englewood Cliffs, NJ: Prentice Hall, 1991); and

R. Glaser, "Expertise and assessment," In eds. M.C. Wittrock and E.L. Baker, *Testing and cognition,* (Englewood Cliffs, NJ: Prentice Hall, 1991).

A third major issue posed is whether literacy can be assessed independently, without requiring contextual knowledge. This issue is given greater consideration in Chapter 3, but Venezky concludes that while "lower level reading skills can be evaluated independently of specific curricular areas, higher level literacy skills probably cannot. At the highest levels of literacy assessment are those skills that require integration of text-derived information with information obtained previously or from other texts in the same task."[276]

Venezky believes "the most important effort for the present project are the various NAEP adult literacy assessments that began with the Young Adult Literacy Survey of 1985."[277] Resnick and Peterson also saw in this experience an opportunity for national assessment of college student learning developers: "We believe that the results of an assessment of this kind could serve as a useful indicator for gauging progress toward Goal 5 and displaying literacy expectations for college graduates to a broad public."[278] As they point out, that procedure

> identified literacy skills as a key element in work readiness. Literacy-related skills are a key determinant of an individual's ability to function both on the job and in the larger society. These skills include not only reading and writing at a basic level, but also the ability to interpret, extract, and apply information from a variety of texts.[279]

"Three general types of tasks were used—prose, document, and quantitative. The categories of tasks are transparent in their meaning or completely distinct. . .[and] were designed to simulate situations where a young adult would be called upon to demonstrate literacy skills."[280]

---

[275] Venezky, op. cit., 5.

[276] Ibid., 30.

[277] Ibid., 8. Venezky describes the Survey as "composed of a series of tasks. Included within the task and the required response were various combinations of reading, writing, listening, and speaking. In the original assessment design, processing demands were classed as knowledge, evaluation, specific information, social interaction, and application. The processing demands of the exercises included in the final test design can be organized in four classes: location or entry of facts, analysis of groups of facts, interpretation, and summarizing."

[278] Resnick and Peterson, op. cit., 18.

[279] Ibid., 18.

(1) Prose tasks required the respondent to demonstrate an understanding of skills associated with interpreting and using information from newspaper articles, poems, and other extended textual material.

(2) Document tasks required the respondent to locate and use information on texts such as labels, charts, paycheck stubs, deposit slips, and order forms.

(3) Quantitative tasks required individuals to perform mathematical operations using figures embedded in a variety of text types.

Resnick and Peterson recommend that an expanded version of the survey should "disaggregate various

> segments of the population, in particular the recent college graduates. This can be done through the type of oral background interview used for the NAEP Young Adult Literacy Assessment. Respondents, however, consistent with the goal of improving the quality of the college educational experience, should be asked to describe the adequacy of their course programs for the demands made upon them in the workplace. The types of occupations in which they are employed should also be identified.[281]

Venezky would also use the Survey as the baseline for a modified version, which he clarifies could assess literacy from two different premises: one where a minimal amount of knowledge is required for a specified task, another where successful analysis of a new text depends on previously acquired knowledge. In text-based instruments, basic comprehension skills involve three types of operations:

(1) Identifying (verification, locating new information),

(2) Analyzing (ordering, predicting outcomes, semantic selection, and argumentation), and

---

[280] Venezky, op. cit., 19.

[281] Resnick and Peterson, op. cit., 24.

(3) Synthesizing (outlining and summarizing).[282]

The other fundamental approach to literacy assessment involves the use of previously acquired knowledge, which must be invoked to accomplish successful integration and analysis of a test text, what may be termed "embedded" [as opposed to separate] assessment. Given this second approach, Venezky demonstrates by example "different levels or depths of analysis upon which college-level assessment might be constructed. These levels are:

(1) Lexical/syntactic (focusing on the specific vocabulary and the syntactic relationships between lexical items),

(2) Propositional (recognizing distinct propositions, and analyzing their style),

(3) Interpretive (questioning what the propositions mean), and

(4) Critical analysis of an almost unlimited scope.[283]

The first two of these levels generally do not require prior knowledge, but the interpretive and critical levels do; Venezky suggests that such levels should permit access to background documents during the assessment, and thus in part rely on further skills such as information retrieval, rapid skimming ability and a high rate of silent reading. One caveat about such a plan was offered by Margaret A. Miller who emphasized that care must be taken in constructing tasks not properly sensitive to the cognitive research:

> The paper does contain a curious suggestion that reading ability might be determined if students were to both read and listen to "equivalent" passages and their understanding of the information assessed. The difference between comprehension of the two types of input would then be an indicator of teachability," by which he seems to mean reading competency. Besides the problems with this approach mentioned in the paper (i.e., text difficulty and intelligence complicate the comparison), aural processing differs in kind from literate processing: material needs to be organized

---

[282] These skills echo the Delphi structure and also that proposed by Banta. These operations are demonstrated by commentary on sample items by Venezky on pp. 12-21 of his paper.

[283] Venezky, op. cit., 26.

> differently in each, and the skills necessary to process each kind of input are different. So comparisons are not only likely to be complicated; they also risk falling into a fundamental category mistake.[284]

In sum, though NAEP uses a functional definition of literacy, Venezky recommends the cognitive approach for higher level functioning of college students.

> Although any cognitive model of literacy competence will change over time, it is probably more reasonable to acknowledge this possibility and nevertheless attempt to define levels of expertise based upon human information processing, rather than to anchor literacy assessment on current demands of societal functioning. Therefore to be compatible with current directions in cognitive assessment, it is recommended that literacy assessment be based upon criterion referenced measurement and that it focus upon those skills that from the best knowledge available we assume are required for expert behavior in literacy.[285]

Venezky believes the ability to evaluate and analyze critically are important outcomes of college education, and must be a focus of literacy assessment. Thus not only reading and writing will be necessary skills, but understanding of charts, graphs and other visual representations—whatever is necessary to find or generate texts that will be the basis for the more extended tasks of identifying, analyzing, and synthesizing. One essential element, he concludes, of such sophisticated analysis is other knowledge (possibly previously acquired) with which test texts must be integrated. This material can be provided as background during the test, with each student working in his or her major area.[286]

Ronald K. Hambleton, from the University of Massachusetts at Amherst, goes further even than Venezky, Dunbar, and Resnick/Peterson in his estimation of the lessons to be gleaned from NAEP.

> For the record, I believe NAEP, generally, has been well-received by policymakers and educators, though, because students, schools, districts, and states are not specifically identified in the NAEP, it is probably viewed by those participating as a

---

[284] Miller on Venezky.

[285] Venezky, op. cit., 29.

[286] Ibid., 30.

> low-stakes assessment. But perhaps there is a message here as well. To paraphrase a comment I heard recently, the more high stakes a testing program is, the less valid the results. If you want valid information about what students have learned, use low-stakes tests, which schools don't view as important! Of course, the problem remains to ensure students are motivated to show what they know and can do.[287]

He would adopt for the national assessment of college student learning the general NAEP framework, preserving what he considers its most constructive aspects:

(1) National curriculum committees to identify the relevant competencies to assess,

(2) Measurement committees to concern themselves with questions of valid assessments and scaling,

(3) Complex sampling designs for selecting participants, and

(4) Thoughtful and comprehensive reporting methods, which have been functioning successfully for over 20 years.[288]

Writing Assessment to Probe Effective Communication

Another advisor who confined his analysis to one particular arena or set of skills was Ed White, who provided an in-depth look at the assessment of writing, perhaps the only ability set that could feasibly capture the Goal 5 reference to effective communication. Writing, defines White, "encompasses a wide range of skills, from the mechanics of punctuation and spelling to the systematic or even creative development of ideas. The high order skills of communication necessarily involve critical thinking and problem solving, and an assessment of the writing of advanced students should focus on these high order skills."[289] Thus writing at a certain level manifests communicating effectively, and provides the vehicle for demonstrating critical thinking and problem solving.

---

[287] Hambleton on Dunbar.

[288] Ibid.

[289] White, op. cit., 1.

White emphasizes a major distinction between *learning writing* (what is referred to as an imitative and socializing skill) and *mastering the art of writing well* in the service of critical thinking, an individualizing skill.

> When we develop arguments, conduct research, or solve problems, such imitation is not only insufficient, but it defeats the purpose; we must think for ourselves, as individuals, if we are to write well. There is no question about the need for these individualizing skills as we consider writing as part of education for the workplace or for citizenship. The information society of the future requires workers and citizens who have learned how to solve problems, to evaluate evidence, to come up with new ideas or new approaches to old ideas.[290]

"The imitative and socializing skills may have some slight role to play, but our principal concern must be upon the ability of students to create discourse as a crucial part of the discovery, thinking, and evaluating process. We cannot rest content with a passive test of passive skills, but must come up with some way to assess actively the active thinking of students as they write."[291] Boehm concurs:

> For a good while now, writing teachers, and, increasingly, teachers in other disciplines, have maintained that most writing, and especially essay writing, unavoidably, involves critical thinking. They argue that to write essays *is* to think critically, that the process of writing *is* critical thinking. Look at what's involved in that process: selecting a focus, finding a way to hold it still, generating ideas about it, choosing from among them, choosing words, ordering them, rejecting, keeping, reconsidering, polishing. Unpack composing an essay. . .look at what mental work doing it requires—and you find it involves taking things apart, seeing relationships, making connections, using judgement, reasoning, evaluating, inferring, and so on—elements, each of them, surely, in any credible definition of critical thinking.[292]

---

[290] Ibid., 3.

[291] Ibid., 4.

[292] Boehm on White. Boehm adds that this "important point [is] substantiated by a considerable amount of the scholarship generated by the writing across the curriculum movement over the past fifteen years:

John Bean et al., "Microtheme Strategies for Developing Cognitive Skills," in *New Directions for Teaching and Learning*, C. W. Griffin, ed., (Place: Jossey-Bass, 1982).

To fully acknowledge and appreciate the critical thinking that is thus wrapped up in writing, insists White, it must be assessed not only as a product, but as an active process. This distinction will be fairly considered in Chapter 3, but White defines the sort of writing he means through the lens of how it will be judged and assessed.

> Thus the process definition of writing, as well as the individualizing function of writing, should lie behind whatever assessment may evolve from the present quest. Process evaluation argues for complex judgments of competence based on more than the correctness of the product. The process model sees writing as a series of overlapping activities, all of which have to do with critical thinking and problem solving: invention and rewriting, drafting, refining and rethinking, connecting, revising, and (finally) editing. The assessor cannot rest content with passive testing of such matters as editing skills or neatness or even achieved organization of writing. Evaluation of writing as process needs to view drafts and revision, as clues to the thinking and writing that go on as the drafts become more and more finished.[293]

The Call for Necessary Research

While NCES stressed to advisors that, difficulties notwithstanding, the workshop process would try to explore what sort of national assessment of college student learning process might be feasible now, some scholars insisted that [as discussed earlier in Chapter 1] in essence, "you can't get there from here." Almost all believed that—in an ideal world magically insured against problems of funding and politics—a very solid critical thinking assessment could be developed. But a solid majority also urged that the pragmatic course be taken, given the impetus and apparent willingness to develop something now. A few,

---

Ann E. Berthoff, "Towards a Pedagogy of Knowing," *Freshman English News*, Spring, 1978.

Janet Emig, "Writing as a Mode of Learning," *College Composition and Communication*, March, 1977.

Toby Fulwiler, "Writing: An Act of Cognition," in *New Directions for Teaching and Learning*. See above.

J.N. Hayes et al., *The Writer's Mind: Writing as a Mode of Learning*, (Urbana, Il: National Council for Teachers of English, 1983).

Judith A. Langer and Arthur N. Applebee, *How Writing Shapes Thinking*, NCTE Research Report No. 22, 1987.

[293] White, op. cit. 10.

however, urged that a more elaborate process of evaluation was needed, suggesting that current experience, and the conclusions to be drawn from extant data, were not solid enough to support a selection of skills and abilities for a working version of a critical thinking assessment. These advisors tended to point to specific difficulties and deficiencies, and to suggest how to deal with them.

Fitting national assessment of college student learning to American Higher Education

James Ratcliff is another advisor who, like Morante and Banta, is working in the trenches of postsecondary assessment. He echoes Lenth's belief that—for any national assessment of college student learning to work—it must gain acceptance among the current postsecondary infrastructure. Ratcliff's conviction is informed by his experiences in assessment within that system, based at the National Center for Postsecondary Teaching, Learning and Assessment at Penn State. He is particularly sensitive to the need for faculty to come to feel invested in any new process, and he believes that—since for a national program to succeed it must be instituted at a majority of institutions in the country—"The national effort should build on the credibility and accomplishments of these programs rather than to be duplicative and ancillary to them."[294]

This caveat leads him to propose a dual approach: "For the short-term, protocols could be developed to monitor selected course syllabi and examinations to determine the extent to which they encourage the development of communications, critical thinking and problem-solving abilities. A longer range goal would be the adoption by states and institutions of a national assessment program."[295]

Ratcliff sees no short cut to a satisfactory consensus on the target skills and abilities because of the heterogeneity of American institutions. "This variation in learning experiences varies the extent and type of knowledge, skills and abilities [that] students acquire and their performance on test batteries."[296] He believes that the phenomena elicited on most current tests do "not necessarily portray year-to-year dips and rises in graduating student abilities"[297] in any underlying sense of intelligence or thinking ability. "Rather," he believes,

---

[294] Ratcliff, op. cit., abstract.

[295] Ibid., 1. Ratcliff, who directs the Center for the Study of Higher Education at Pennsylvania State University, is also the Co-director at the National Center for Postsecondary Teaching, Learning and Assessment, where, he says, "We are developing such protocols for analysis." [p. 25.]

[296] Ibid., 22.

> it may be simply a function of the courses and subjects they choose to study. When a university offers an undergraduate the opportunity to pick 35 to 45 courses from a curriculum of 3,000 to 5,000 courses to complete the baccalaureate, it is little surprise that different college graduates evidence different levels and types of knowledge, skills and abilities. The challenge to an effective assessment program is to cast its definitions of what constitutes clear communications, critical thinking and problem solving sufficiently broadly to capture the full range of learning associated with these terms.[298]

Ratcliff suggests a research program to this end but, like a number of his colleagues, cautions against "accepting current critical thinking instruments such as the Cornell Critical Thinking Test or the Watson-Glaser Critical Thinking Appraisal, in part because most students demonstrate a high correlation between results on these and on more standardized measures like the ACT, SAT and GRE examinations."[299] A more focused approach—which has been taken at the National Center—is to look at parts or sub-elements of such instruments that correlate with the target abilities. "Students in the Differential Coursework Patterns (DCP) Project. . .showed significant improvement on the Analytic Reasoning (ARE) and Logical Reasoning (LR) item-types of the Graduate Record Examination."[300]

Ratcliff restates the logical premise of his research as it applies to the national assessment of college student learning:

---

[297] Ibid., 22.

[298] Ibid., 22.

[299] Ibid., 23.

[300] James L. Ratcliff, *Development and testing of a cluster-analytic model for identifying coursework patterns associated with general learned abilities of college students: Final report*, (University Park, PA: Center for the Study of Higher Education, Pennsylvania State University, 1990).

> To what extent do these GRE item-types represent critical thinking or higher order reasoning skills? We need further research to identify the types of knowledge and cognitive abilities required to answer these type questions.[301]

This necessary first step of correlating types of questions to certain specific abilities then provides a foundation for educators to ask the larger questions addressed by the NCES exercise:

> To what extent do they, for example, suffice a college's need (or a nation's need) to measure the development of critical thinking abilities among its students?[302]

Ratcliff's research suggests that it is important to approach the identification of skills and abilities pragmatically, looking for their correlation with specific courses, student experiences, and perhaps outcomes on instruments related to the analysis of specific item-types:

> Using the methodology and model from the Differential Coursework Patterns Project at the National Center, we could begin now to determine which measures of critical thinking, communications and problem solving best differentiate between appropriate and inappropriate learning environments for students of different ability levels. Rather than become embroiled in debate over what constitutes critical thinking or clear communications, we could begin an investigation of what existing measures overlap each other, which best describe student improvement, and which are most closely aligned with the curriculum of particular institutions or student groups. This second prong of investigation would move us closer to understanding *how* we may use assessment information to improve students' abilities in these key areas articulated in the NEGP objectives.[303]

Ewell and Jones collaborated on a paper which reinforces Dunbar by their conviction that direct assessments of the Objective 5 abilities "are technically complex and will take many years to develop."[304] Thus, while they do not directly advise on how to identify the specific abilities for testing, their research proposal points to a number of what they call "indirect indicators." Conceding that the special nature of such indicators makes them more than one step away from a tool for identifying the Objective 5 abilities, they nonetheless believe that the domain of the assessment should consist of "institutional policies and practices known or presumed to affect the development of higher order skills."[305]

---

[301] Ratcliff, 1991, op. cit., 23.

[302] Ibid., 23.

[303] Ibid., 26.

Thus, rather than a list of skills and abilities that individual students would be tested for, they suggest a system of evaluating the overall situation in American colleges. The Objective 5 abilities would be approached by analyzing the requirements institutions made of their graduating students, surveying "good practices" as they are manifested in instruction, and eliciting a profile of student behavior. To suggest the time-frame they think realistic to develop a fully functioning national assessment of college student learning, they predict even this first stage they outline would take three to four years before useful, integrated information could begin to provide a solid enough base on which to begin to construct the more permanent system.

Don Rock provides a view of the national assessment of college student learning process from deep inside the world of instrument construction and analysis, at the Educational Testing Service (ETS). His viewpoint reflects experience with a number of instruments and the processes that led to their development. He suggests that "*expert* judges from many different perspectives be used in the goal setting procedure but be given different tasks depending on their expertise."[306] Such talent could then undertake an exercise designed to identify and define

> the knowledge, skills and abilities that are necessary to perform successfully at each of the performance levels in each of the content areas. It is argued here that individuals from industry can play a significant role here since they should be aware of both the level and kind of knowledge, skills, and attributes necessary to successfully make the transition from college to successful performance in their jobs. Educators and test specialists can then develop and/or match presently available items to the list of desired skills and skill levels.

---

[304] Ewell and Jones, op. cit., Abstract.

[305] Ibid., Abstract.

[306] Rock, op. cit., 16.

Then the representatives from industry and government can rate the items in the pool with respect to their relevance to the desired skills. Then, using those items that have been judged to be relevant, college teachers who have been involved in the whole Knowledges, Skills and Attributes Process and are familiar with both typical college student performance as well as testing, can make the judgements about expected student item performance for the various criterion referenced levels. This strategy is consistent with the procedures used by industrial psychologists in developing tests to measure skills required in industry.[307]

Rock's plan begins to build a bridge between two worlds that are often speaking truly different languages. Tenopyr injects a note of realism about how far such a plan would need to go, and her concerns also suggest some of the issues that might arise in trying to extrapolate from the Alverno experience. She classifies the "difficulties in doing research on the relationship between college achievement and later success in life:

1. The difference in difficulty of various curricula both within colleges and across colleges.

2. The overall differences in grading practices among colleges.

3. The differences among employers in selection practices relative to preferences for college attended and grades received.

4. The differences, relative to employer size, regarding the way positions for college graduates are filled.

5. The problems of correlating grades with salaries.

6. The effect of self-presentation skills on success in many endeavors in life.

7. The lack of systemic promotion procedures in most companies.

8. The effects of race, nationality and gender on measures of job success.[308]

---

[307] Ibid., 16

[308] Tenopyr on Cappelli.

Ultimately, what she and others are cautious about are attempts to translate one body of experience into a different context. The success of any such translation always comes back to the frame of reference, and this frame must be approached in terms of standards, which is the issue of Chapter 3.

3. STANDARDS AND OTHER IMPORTANT MEASUREMENT ISSUES: SIX IMPORTANT QUESTIONS

Included in the original charge to authors was the question, What Performance Standards Should be Established? Most of them found it difficult to provide a straightforward response. In part, this difficulty seemed to be one of putting the cart before the horse: in working through the conception of their particular version of a national assessment of college student learning, most discovered that a certain sequence for the process of development was emerging, and that trying to take a subsequent step before earlier, foundation issues had been clarified was to prove slippery. Nonetheless, the dialogue between authors and reviewers, and subsequently among discussants at the workshop, was rich with discovery because—in the attempt to talk about standards—most of the major dilemmas were revealed, and often experienced in heated debate.

This debate and discussion proved to be an extension of many of the central concerns described in Chapter 1. There the issue was, What does the national assessment of college student learning mean? Here, scholars contend with the crucial corollary: What will the structure, results, and standards of a national assessment of college student learning actually say about those tested (and how reliably and validly say it) and how will that information be interpreted? As the NCES workshop participants fenced over definitions and important frames of reference and issues of validity, this larger question devolved into a number of more pragmatic, more specific ones. The *terms* often used to frame these questions recur throughout the NCES deliberations: in the authors' briefing paper, the original papers by the authors, the commentaries, and the discussions, summaries, and remarks recorded during the workshop. While fairly common terms in this field, they need to be defined for the purposes of discussion, as various scholars used them with different twists and intent. So first, some working definitions:

Domain generally means one of two things. Some used it synonymously with "category of abilities." Such a use, where relevant, was referred to the previous chapter. For the purposes of this chapter, domain means "category of knowledge that underlies an item, section, or entire assessment." The main issue to be considered is whether specific subject-matter (content) domains are essential to a higher-level cognitive assessment of critical thinking.

Transfer, in this particular discussion, refers to how large a significance can be attributed to the results/outcomes of a specific test, or to the results obtained within a specific domain. A national assessment of college student learning that effectively probed Objective 5 abilities, *with respect to transfer*, would assess college student learning for results that were widely applicable beyond the specific assessment situation itself. Inferential judgements could then be made with respect to both ability in areas of knowledge or contexts not actually used in the assessment, and/or to transfer to the world (of work) more generally.

Context in the national assessment of college student learning has two basic meanings. The first, more specific, meaning circles back to the definition of domain. Assessments avail themselves of a context of assumed knowledge. This knowledge can be either

- generally assumed awareness related to the age and educational experience, of the subject,[309]

- specific knowledge which is provided the subject to manipulate in demonstrating the necessary critical thinking skills, or

- knowledge of a particular subject area presumed to have been mastered in the pursuit of a particular subject area/course of study, or curricular program.

An entirely different use of context refers to the overall political, social, and educational context in which the national assessment of college student learning is considered. These issues were discussed extensively in the first chapter on the meaning of a national assessment of college student learning, where certain scholars felt it crucial to frame their approach with an analysis of premises and goals. The most prominent such issue seemed to be that of Accountability vs. Improvement; that is, in its starkest terms, Should the national assessment of college student learning be a single instrument that will lend itself to political and institutional scorekeeping? Or should a national assessment of college student learning be an umbrella to embrace a number of approaches, systems, and feedback loops all of which subserved a larger goal: to enhance and inform the relationship between the outcomes being reported and the educational processes and systems that presumably underlie them.

---

[309] And—as many in the workshop stressed—the cultural background. This issue will be taken up later under Question 5, but it should be understood that a long history and literature has developed over the question of cultural bias. Suffice to say here, the theme was a recurrent one throughout the NCES process, and could be mentioned—not only with reference to all three possible knowledge contexts—but to almost any element of the NACSL under consideration.

131A

Proficiency levels refers to the question of explicit standards by which the scores/outcomes of a national assessment of college student learning would be evaluated and compared. Rather than a single scale and universal set of criteria, multiple standards could be established to subdivide test subjects by various background criteria, as well as to reduce the pejorative difficulties that reporting exclusively by institution could create.

Carving the national assessment of college student learning process up into neatly defined and distinct issues and pieces might facilitate discussion, but the authors' papers and the workshop itself illustrated that no such clear-cut divisions can be maintained.[310] Standards may be clearly one animal if by standards you mean "numerical scores and their transposition onto a social scale of value." However, when the term standards is taken—as it often was—to represent the underlying component elements that should be revealed in demonstrating the target skills, it becomes another animal altogether (and incidentally, an animal that wants to graze on the terrain covered in Chapter 2 on identifying which skills to target). Thus, this chapter follows up on some of the material already presented, but from a different point of view. Identifying skills in relation to the substantive concept of critical thinking (Chapter 2) is one thing. Trying to envision how those skills could be properly assessed is quite another, because a number of other issues are forced out into the open:

- Is the assessment coherent in pointing toward a goal; e.g., actually testing the possession of abilities rather than the ability to perform on a test?

- Is Goal 5 one goal or several; and if several, are they compatible? If one of the goals (indeed in Chapter 1 several authors argued that it should be the predominant goal) is to inform and improve instruction, can the process ever unify the many political and infrastructural elements involved?

- Can a unified assessment that tries to do so much be valid, in a strict social science sense, given the several fundamentally different types of colleges and universities, and the enormous diversity of those to be tested?

---

[310] Again, a reminder. The authors were charged with the overall conception of an NACSL development process, and asked explicitly to advise on a number of specific questions. Three of those questions have been separated out from the current analysis, for later consideration. Thus, Chapter 3 embraces the remainder of relevant issues and material *except for* these three issues: when in their career should students be tested, how can they be motivated, and what sort of instrument should be developed.

- Even if an academic model of how to construct a national assessment of college student learning is postulated, are the background research and premises on which it rests sufficient to fortify it against the inevitable criticism it will encounter—given the several different constituencies affected, and the various, possibly paradoxical, purposes it hopes to serve—when implemented?

Ultimately all authors and subsequently the workshop participants seemed to be struggling with such fundamental questions. While they were occasionally addressed directly and explicitly, often they became the context for debating the issue of ***Standards***. Standards, in turn, led the authors and workshop participants to frame questions about content domain, general vs. subject-specific skills testing and how they transfer, levels of proficiency, and the social and political motivations and uses of the assessment. These are the themes that recur throughout Chapter 3. The discussion is structured as follows:

(a) Are standards inherent to the elemental task of defining a national assessment of college student learning?

(b) What is the historical context for deriving and implementing standards?

(c) How would standards in a proposed national assessment of college student learning relate to the overall charge of Goal 5—do they "transfer" beyond the academic setting?

(d) More specifically, must a national assessment of college student learning test subject-specific content domains in order to generate robust and reliable conclusions about transfer?

(e) Is a single set of standards reasonable, or even possible, given the diversity of institutions and of those to be assessed?

(f) How might the foregoing suggest a particular "brand" of assessment?

a. Are standards inherent to the elemental task of defining a national assessment of college student learning?

For the critical thinking movement, the question of standards is inextricable from the overall planning and development of any national assessment of college student learning. Paul and Nosich adapted the canon of critical thinking precepts developed at the Center for Critical Thinking into a list of 21 Objectives they commend to national assessment of college student learning developers. Depending on the focus, their list could be considered for its implications for both Chapters 2 and 3. That is, their description of what critical thinking skills *are* amounts to a definition of those skills, whereas when they say what such skills *should accomplish*, they are providing a value judgement which—if accepted—becomes a template for applying standards to the process.

Before addressing the question of explicit standards to be derived and applied to a finished instrument, however, they argue, there is a deeper level. Critical thinking is an activity that cannot be conceptualized without the most fundamental of standards. "Critical thinking is based on the art of monitoring one's thinking with standards implicit in the universal structure of thought." All thinking, ergo, invokes them: "the use of these standards with respect to the structure of thought is implicit in intellectual history from Socrates through Einstein."[311]

As often happened during the NCES exercise, scholars were aware of the irony of themselves bringing less than the most acute kinds of critical thinking to the problem-solving exercise at hand: to wit, developing the best possible national assessment of college student learning, in light of Goal 5 and their own experience. They were, in a way, 'through the looking glass' with the obligation to exemplify critical thinking in their deliberations. Thus, at this first step, the national assessment of college student learning development process "should be based on clear concepts and have well-thought-out, rationally articulated goals, criteria, and standards,"[312] insist Paul and Nosich. In this spirit, the standards that could be established as benchmarks of comparative performance will be termed *achievement* standards and the implicit values the *inherent* standards.

---

[311] Paul and Nosich, op. cit., 6.

[312] Ibid., 6.

Since critical thinking thus seems to offer "intellectual standards [that] apply to thinking in every subject," (including the activity of developing a national assessment of college student learning) the authors offer a menu which exemplifies this sort of universal distinction between good critical thinking—in which the universal intellectual standards are implicit—and less than good thinking, irrespective of the domain, content, or any other context in which the thinking activity may be framed for evaluation.

*"**Intellectual Standards That Apply to Thinking in Every Subject**"*—what, in the suggested terminology, are inherent standards—they describe by distinguishing between

thinking that is:		**thinking that is:**
Clear	vs	Unclear
Precise	vs	Imprecise
Specific	vs	Vague
Accurate	vs	Inaccurate
Relevant	vs	Irrelevant
Plausible	vs	Implausible
Consistent	vs	Inconsistent
Logical	vs	Illogical
Deep	vs	Superficial
Broad	vs	Narrow
Complete	vs	Incomplete
Significant	vs	Trivial
Adequate *(for its purpose)*	vs	Inadequate
Fair	vs	Biased or One-Sided[313]

During the workshop, Paul emphasized this point:

> And this is integral to our understanding of what critical thinking is. . .We're talking about the kinds of intellectual standards that students should come out with at the end of their college career. It is my observation that if you ask most graduating seniors, "What intellectual standards have you learned, that you now hold your thinking

---

[313] Ibid., 16. [This is the chart referred to early in Chapter 2.]

> responsible for?" you would find that students would draw a blank. That is, present instruction does not call attention to intellectual standards. It tends to be heavily focused on content, and the re-iteration of content in lectures and textbooks. So I think there's a substantial problem here. And if you understand critical thinking is connected with intellectual standards, you see it in a somewhat different way, and I think a richer way.[314]

There is an unlimited nest of chinese boxes where this concept might apply, including the process of fostering critical thinking, regardless of whether in an explicit course of critical thinking instruction or in the context of another subject. Thus, such inherent standards would apply to any and all discussions of the national assessment of college student learning, even those where an instrument was opposed in favor of a process predicated on better informing instruction. Indeed, say Paul and Nosich, "the process of learning to teach so as to foster critical thinking is the very process by means of which one establishes intellectual standards for assessing thinking, and, by extension, for assessing instruction itself."[315]

These inherent standards, then, have been virtually elevated to principles. They need not be debated, nor criteria-referenced, nor adapted (in any major way) to the context of the test nor for different test-taking audiences. They are the essence of critical thinking, and therefore provide a lens through which to view not only any model and its component parts that might be proposed by national assessment of college student learning developers, but also the very process by which such results have been arrived at.

Other scholars, for more pragmatic reasons, shared a similar premise: that embedded in the very conception of a national assessment of college student learning should be some "higher goal" or implicit standard. (Not to confuse the chicken with the egg, they are talking about an inherently high set of *achievement* standards.) Venezky, for example, in drawing lessons from the assessment of literacy, believes that "to be compatible with the basic directions of critical thinking and communications research, it is suggested that literacy assessment be built around a definition of human expertise."[316] Shoot higher than a functional definition of standards, he advises, and use the broad base of cognitive research to develop inherently and comparatively high standards.

---

[314] Richard Paul in Open Session.

[315] Ibid., 16.

[316] Venezky, op. cit., 2.

Lenth urged his colleagues to consider the value of such an approach for another, practical, reason. His survey of recent assessments in the American postsecondary universe indicates that, when such efforts are mounted, the political and structural temptation is to shoot low. "Lacking the necessary means for direct cognitive assessment, and disinclined to do so for a variety of reasons, most states attempted to improve undergraduate education by pushing on the bulging middle, rather than by pulling from the top by specifying higher expectations and attempting to assess higher levels of performance."[317]

Of course Lenth's survey was confined to the postsecondary universe. His reviewer Swanson noted that

> the efforts in some states could best be described as 'pushing from the bottom' where the intent is to improve secondary (or earlier) education as a means of better preparing students for the college experience. My point here is that a national assessment program will have to deal with deciding where educational interventions, if any, should be attempted and the experience in at least some states would indicate that effort may not be best directed to the improvement of undergraduate education, but rather to much earlier educational levels.[318]

Standards, obviously, will not materialize out of the political thin air of any national assessment of college student learning if they do not derive from, relate to, and become systematized throughout the K-12 system. Swanson states eloquently the obvious:

> The national goal and the related objective lead me to surmise that in order to substantially increase the proportion of college graduates who demonstrate the skills in question, assessment would have to take place at several educational levels to determine when such skills are developed so that timely corrective action could be applied in order to affect the outcome relevant to the national goal. This may involve assessment in high school or even before. The core issue here is that we may have to assess prior to college in order to have any realistic chance of meeting objective five of goal five.[319]

---

[317] Lenth, op. cit., 7.

[318] Swanson on Lenth.

[319] Swanson on Morante. Again, the questions of when in their careers students might best be tested and how a baseline might be established and used, have been postponed for future study.

But Lenth and others during the workshop emphasized two stark realities: first, a national assessment of college student learning is what is on this particular table at the moment, and that a sort of "trickle down" effect is inevitable. He uses this reasoning to urge national assessment of college student learning developers to set inherently high standards, and thereby avoid any seemingly direct encroachment on the sort of testing for more functional abilities that many state educators consider their province.

Lenth has a second, more urgent, motive for supporting high standards: a somewhat hardened view of the political art of compromise as it may devalue the national assessment of college student learning undertaking. He notes that "the language and processes of assessment [are] now part of the policy environment" at all educational levels, and wonders whether "both the idea and the wording"[320] of Goal 5 (particularly Objective 5) might not simply have been imported from the ongoing dialogue about assessments of elementary and secondary education. Lenth looks apprehensively at the postsecondary dialogue and raises the concern that Goal 5 could experience the same bowdlerization he sees in the "nation's first 'report card,' the *National Education Goals Report: 1991*," where the language describing these issues in the Romer committees' reports and the earlier Goals discussions is nowhere to be found. Lenth notes that "later on in this document the description of the proposed assessment has been reduced to the words "*a national assessment of college students' thinking, communication, and problem-solving skills.*" The vision of enhancement, 'advanced' abilities, 'critical' thinking, and 'effective' communication has been taken from the language. Have we already begun to trivialize the challenge and the vision of postsecondary assessment?"[321] he asks, fearful that by compromising the vision of inherent standards, developers could set the stage for lower achievement standards. The answer isn't rhetorical as much as

> nearly unavoidable without a clear and widely-communicated statement at the outset that defines the purposes and intended uses of the assessment process. Such a statement is a fundamental component of academic policy and planning, and needs to be set in a policy framework that defines both expectations and means for policy implementation. I believe that the collective experience of both state-level higher education agencies and institutions provides ample evidence for this position,

---

[320] Lenth, op. cit., 21.

[321] Ibid., 21-22, citing the National Education Goals Panel, *The National Education Goals Report: 1991 — Building a Nation of Learners*, Washington, D.C., September 30, 1991, p. 21, 192.

> illustrating cases in which assessment has failed and been trivialized because of uncertainty over purpose and intended uses, as well as instances where reaching agreement over academic goals and purposes, and how assessment can contribute to these, was the most formative and fundamental part of the process.[322]

States have to be motivated to support a national assessment of college student learning, and Lenth thinks the highest goals and achievement standards will be most attractive to that group of users. His warning is forceful: just because the debate about standards may unleash a number of positions that reflect the perceived threat to some aspects of the infrastructure, don't lose sight of how integral clear and high standards are to the vision of a national assessment of college student learning that he believes will ultimately appeal to the audience which (for it to work) has to accept it—the educators in the states.

> The greatest risk to this [national assessment of college student learning] effort would be in the trivialization of the objectives and the agenda. In order for participation in the assessment process to be worthwhile, institutions and state leaders would need to get something back. That something, I believe, involves a clearer definition of expectations for higher order ability levels of college graduates, the development of sophisticated methods to assess these, and the eventual development of benchmark data and a research base to be applied to improving education in these areas.[323]

In Chapter 1, Dunbar raised concern about some of the underlying premises of undertaking a national assessment of college student learning. He believes one of the most crucial of these involves the magnitude of the gain called for in Goal 5, and how the implicit (inherent) standards of such a test could complicate the pursuit of that goal. Flat out, he warns developers, "the most challenging task facing NCES given the language of *America 2000* is the development of achievement standards for college students in the United States."[324] The language he is referring to comes from the *America 2000* goals statement, with "its focus on gain in critical thinking, communication, and problem-solving skills, gain of *some unspecified* magnitude that in the judgment of, say, some *unspecified* blue-ribbon panel, is an amount that will prepare an educated citizenry to be competitive in a global economy of dwindling resources."[325]

---

[322] Ibid., 22.

[323] Ibid., 24.

[324] Dunbar, op. cit., 14.

Dunbar's concern has to do with the naked fact that the national assessment of college student learning emanates from the political realm, and points directly toward a program of academic reform that will be judged by its social utility. Taken literally, Goal 5 could become the foundation for an enormous expenditure of resources and a re-direction of American higher (and, inevitably, lower) education, and could foster *measurable gains* in critical thinking, by achievement standards yet to be established, and then *fail* in its ultimate task, because the social utility to which all of this energy would be directed may not be demonstrably achieved. It is not believed that the explanation "Well, we didn't control for the global economy," will be adequate. As was forcefully emphasized by a number of authors, and clarified in Chapter 1, it is most unlikely that this particular circle will be closed. That is, unless the national assessment of college student learning takes a dramatic turn in a direction that would be revolutionary in American assessment, the best that can be done is to focus on fostering critical thinking skills that we *hope* or believe or (based on what many would characterize as insufficient, at best "soft" evidence) postulate will achieve the "citizenship" and "globally competitive" aspects of Goal 5.

Another author was keenly aware of the recent history of assessments gone (at least partly) awry. Rock's experience with ETS provides him a perspective on how these things actually run, once they are gassed up and let loose on the track.[326] Like Lenth, he believes that the setting of standards—for Lenth they could embody the purposes and goals of the assessment—is a crucial element in the mix, for several reasons:

(1) Setting realistic goals should be the first step in bringing about change in any complex delivery system. The goals must be realistic or the resulting frustration on the part of both teacher and student will only make the situation worse.

(2) [Thus] the setting of performance standards in assessment tasks may help with motivational problems.

---

[325] Ibid., 4.

[326] As an example, the NAEP mathematics assessment inaugurated in 1990 for the 4th, 8th, and 12th grades, with the imposition of standards by the National Association Governing Board. Rock's view of this experience ("Much can be learned from this ambitious effort about the difficulties in setting standards") will be discussed later in Chapter Three.

> (3) Performance standards can provide some diagnostic information concerning educational deficits, and if they (the standards) are accompanied by specific behavioral descriptions they could give some direction as to how to remedy those deficits.[327]

Following up on Rock's view about the specific benefits of standards, Paul and Nosich report that standards have been found to be "more useful if they are made explicit to the students who are taking the test because. . .

> they can then see that there are standards, that the standards are not arbitrary ones, and that understanding the standards gives them an insight into what good critical thinking is,

to those doing the assessing because. . .

> in addition to those reasons, it fosters both a uniformity in grading and a strong correlation between the grade and the skills being graded.. . . Thus, making standards explicit promotes both the reliability and the validity of the assessment-vehicle,[328]

and finally it benefits classroom teachers because. . .

> such standards can readily be built into classroom instruction. The standards, after all, are those implicit in teaching for higher order thinking skills; they are therefore invaluable both for teachers to use explicitly with their classes and—an essential feature of critical-thinking-internalized—for students to learn to use as part of assessing themselves.[329]

Thus do inherent standards apply to almost every imaginable aspect of the national assessment of college student learning. They will impact every player, and actually be implicit at every stage of development. For starters, they should be high (in the argot of critical thinking, "richly conceived"), woven into the very fabric of intellectual assumptions on which the test will be constructed, defined from a cognitive perspective (and towards the higher end of that realm) and formulated explicitly and communicated clearly to all involved. Given how central an element both inherent and achievement standards are to assessment, how have they been deployed in the past?

---

[327] Rock, op. cit., 13.

[328] Paul and Nosich, op. cit., 16.

[329] Ibid., 16.

b. What is the historical context for deriving and implementing standards?

Several authors addressed this question with antennae tuned for lessons that have—or should have—been learned. Ratcliff describes a pattern that recurs throughout American educational history. Beginning at Harvard, and down through the establishment of land-grant colleges and then special emphasis institutions—for women, native Americans and blacks—universities of some quality generally *preceded* the creation of secondary and primary school systems that could prepare students adequately to enter them. "This curious historical phenomena consistently has placed higher education in the position of judging the qualifications of the students it admits, thereby articulating academic standards for college preparatory and secondary education in the process."[330] Once secondary school systems had been established, the pattern continued in the form of explicit admission standards (as colleges perceived that secondary preparation was often inadequate) and standardized tests to evaluate whether students were meeting them. Thus was born, from the need to protect colleges against a flood of ill-prepared applicants, the College Entrance Examination Board (CEEB).

Ratcliff emphasizes that "the CEEB program was *not* imposed on colleges and universities by state or federal government. Rather, the need was widely seen by the college leaders of the day, thereby enforcing its widespread acceptance and success in higher education."[331] The CEEB accomplished its avowed purpose and "*did* dictate standards for secondary education and for student performance in key areas of knowledge, skill and ability."[332]

But while some uniformity of quality was eventually assured by precollegiate standardized testing, the fabled American melting pot was beginning to stir. Ratcliff explains that "the rapid rise in immigrants during the first two decades of this century" introduced an element that was to become endemic, ultimately characteristic of American education; to wit, the striking diversity of both the educational backgrounds and [therefor] the needs of postsecondary applicants. "Since 1980, there have been 6 million new immigrants to the United States. . .the diversity of cultures and educational backgrounds of college students is expanding rapidly."[333]

---

[330] Ratcliff, op. cit., 3.

[331] Ratcliff, op. cit., 4; citing Brubacher, J.S. and W. Rudy, *Higher Education in Transition: A History of American Colleges and Universities*, 1636-1976. 3rd ed. New York: Harper & Row, New York, 1976).

[332] Ibid., 4.

The significance of this diversification, argues Ratcliff, is a latterday echo of the situation that led to the CEEB. Because the "quality, relevance and recency of [this group's] secondary education" is less uniform and ascertainable," the American educational system—especially at the interface between secondary and postsecondary—needs "to set clear standards for the articulation of secondary, precollegiate and higher education."[334] However the very impetus for this need—the diversity of the postsecondary institutions and their curricula, and of the backgrounds of their students—poses a central dilemma for the national assessment of college student learning: can an assessment embody standards that harnesses this diversity to Goal 5? A large question, which will be taken up in some detail later in the chapter. But a good foreword to that discussion is the perspective provided by Morante, whose excellent history of statewide assessment in New Jersey during the last decade comes from the inside, as he served as Director of the College Outcomes Evaluation Program (COEP).

The New Jersey experience reported by Morante actually covers two different testing eras, and the statewide test which most closely augurs the national assessment of college student learning was administered only twice, and then fell victim to a change of administrations and the political axe. However the General Intellectual Skills (GIS) test *was* administered to samples of students at most public institutions in both 1990 and 1991, and Morante's in-depth narrative of the New Jersey experience illustrates some of the practical problems encountered there.

Questions of standards, both inherent and achievement, recurred throughout the New Jersey process, from initial definitions, to framing the original plan, to formulating a strategy for test development, to deciding on how to score and report results, and—most dramatically—to the social and political impacts the assessment process was seen to have from beginning to end. As Morante describes it, "the first administration of the GIS

---

[333] Ibid., 6.

[334] Ibid., 7.

Assessment was not without trouble." Though little or no faculty resistance was evident at the community colleges or at the universities. . .the faculty union at the state colleges organized a boycott of the test which significantly reduced the number of students tested at most of these institutions."[335] For Morante the reasons were clear. The GIS was a "sophomore test," and therefor, at least implicitly, faculty were being judged on how well their charges could perform after two years of instruction. As he put it,

> A test that assesses students *after* completing a sizeable portion of their college education was perceived as a measure of the effectiveness of the college education they received. This was scary to many in the state's colleges and universities. . . There is little doubt that the fear of a sophomore test was directly related to a fear of being held accountable. Board of Higher Education members publicly proclaimed that accountability as well as using the information to improve student learning, were the reasons for implementing an assessment program.[336]

Beyond simple accountability, the New Jersey planners discovered a number of concerns about such gateway examinations, some of which would not apply to a national test administered to graduating seniors (unless it were a prerequisite to graduation), but others of which were echoed by the participants in the NCES exercise:

- Traditionally under-represented students especially minority students, would be most seriously impacted;
- A gateway test would place the burden of responsibility on students rather than on the faculty or the institution to seek improvement in teaching and learning;
- A strength of the American higher education system is its diversity and any common test would undermine that diversity and result in weakening higher education;
- A statewide [nationwide] test drives the curriculum, *or* what is measured is what is important.[337]

---

[335] Morante, op. cit., 24.

[336] Ibid., 11.

[337] Ibid., 12.

The New Jersey experience reinforced the principle that "If a college awards credit for courses students complete, that institution must be held accountable for ensuring that students are learning."[338] There was no boycott in 1991, reports Morante, but he concludes that "change is not easy in higher education," and emphasizes that "the strong opposition by some administrators, especially college presidents, and at least one faculty union [there are three major organizations in New Jersey] have implications for implementation."[339]

Larson picks up one such among what he sees are a number of "profound possibilities in the assessment movement, related to Goal 5, for substantial improvement in undergraduate instruction."[340] We will need to develop "an awareness of how to approach faculty about changing their orientations toward undergraduate teaching," by fostering the importance of undergraduate teaching and of learning how to learn. He asserts:

> My own conviction is that we have to infuse more widely into undergraduate curricula the acceptance of writing and thinking and critical thinking and problem solving and also problem posing as important elements in teaching processes and learning processes. In order to make that kind of infusion. ..we will have to engage in substantial efforts at faculty development, and helping faculty understand how they can do it; develop confidence that they can do it; and develop the recognition that by doing it they will in fact enhance their teaching—not detract from it.[341]

Greenberg also viewed the assessment environment as a mirror in which society could be perceived. Out where the people are, she reports, "we do not simply enter, muddle through, and go in four years, in the old pattern. We. . .need to pay attention to. . .who these learners are, and how they go through what I call the good revolving door. They come and they go, and they come and they go. And they're stopping in and stopping out, and stopping in, and stopping out." She perceives this pattern as deeply rooted in the social fabric, "a pattern that we need to affirm, that we need to understand. . . Age appears not to be an issue," she believes, when people perceive the use of universities in a kind of cyclical pattern, over their entire lifetime. Assessment, in this view, could become more like asking them the question "How has it worked for you?" says Greenberg. This idea of occasional and periodic interventions suggests that

---

[338] Ibid., 14.

[339] Ibid., 24

[340] Richard Larson in Open Session.

[341] Ibid. Larson adds: "I see in the movement that this conference is initiating a very important chance to make improvements in undergraduate education, sort of along the lines that Ernest Boyer was talking about in his book called *Scholarship Reconsidered*. Maybe we can, through this kind of movement, make scholarship include strength in teaching, and strength in assistance to students' learning."

> We might pull those moments, the moments of assessment, and follow those persons over time, and I think in many ways that's quite affordable and doable, and in fact if we don't do it we will have a very narrow snapshot. And I don't think that's what the intent of the goal structure is. The decade-long 2000 focus, the 2020 focused adventure, which we are now given the opportunity to join. . .[is] really very powerful, [providing] longitudinal looks at how people learn through life and use institutions to do that as they go.[342]

c. How would standards in a proposed National Assessment of College Student Learning relate to the overall charge of Goal 5—do they "transfer" beyond the academic setting?

Among the lessons recent educational assessment efforts can provide is that certain projects get do get proposed, funded, and actually conducted that have no realistic hope of ever achieving—by the strict tests of social science—what the scholars call consequential validity. Dunbar reminds his colleagues that "achievements that are, by their very nature, meaningful to the public in the context of a real-life setting do not lend themselves well to observation and objective measurement."[343] Social scientists struggle with this dilemma, and look for surrogates they hypothesize might generalize to the behavior that is not as directly approachable for measurement. But these models and their premises, according to a recent study by Dunbar and his colleagues, "are prone to charges of contrivance, and the data from them don't generalize well to the criterion situations of real interest."[344]

---

[342] Elinor M. Greenberg in Open Session.

[343] Dunbar, op. cit., 4.

[344] Ibid. 4; see Dunbar, S.B., D. Koretz, and H.D. Hoover, "Quality control in the development and use of performance assessments." *Applied Measurement in Education, 4,* (In press).

Dunbar seems to be saying that, when the smoke is blown away, the real standards (high achievement) or values (inherent) called for by Goal 5 are "whatever will work to achieve the social goal," in essence revisiting a terrain where a number of treacherous pitfalls await the educator who may one day have to rely on sound data and models to defend a program in the political realm. As Dunbar puts it:

> Unfortunately, the use of tests of whatever kind to create an atmosphere hospitable to educational reform is an area of great uncertainty. It is a use of assessment about which measurement specialists know the least vis-a-vis consequential validity. We have no experiments to fall back on, only a history of experiences, largely negative experiences, from the use of tests to drive educational reform in the public schools.[345]

He goes on to describe some of these experiences, such as the minimum-competency testing (MCT) movement of the 1970s in states like Kansas and Florida. While these experiences may not be directly relevant to the national assessment of college student learning (for several reasons: they were gateway tests, the standards were "minimum standards," and the explicit goal was to improve instruction), they did lead to a landmark court case, ***Debra P. v. Turlington***, which unmasked "some of the consequences of basing a decision of enormous social import on a single score,"[346] explains Dunbar, and caused states generally to approach "the task of using tests to make decisions about individual students with much greater attention to questions of consequential validity."[347]

The concept of a gateway test understandably carries dramatic implications for any educational system that is based on progression through explicit levels of achievement. As Morante described in the New Jersey story, the buck stops here, and someone has to be held accountable for the results. Secondly, resources may need to be allocated for remediation, and the entire value structure thereby modified, in effect, because postsecondary institutions then assume the burdens that might have been taken care of at the secondary level. [And, as Dunbar notes here, a gateway concept may serve primarily to reveal a kind of systemic discrimination at all lower levels, anticipating issues of race and class to be taken up later in the chapter under question e.]

---

[345] Ibid., 9.

[346] Ibid., 9.

[347] Ibid., 10.

But the question on the table here is transfer: Are there identifiable skills—which are built around a certain implicit and explicit value structure and can be identified and probed in an assessment—that will actually reveal what those calling for a national assessment of college student learning are after: the objectively measurable enhancement of critical thinking in people who are thereby better prepared to compete globally and to serve as American citizens? In early deliberations about a national assessment of college student learning, nothing like a consensus has been reached to suggest that the test, if administered to seniors, would be a pre-requisite to graduation. Goal 5 calls for a report card, not an explicit gateway, however likely it is that such a benchmark might begin to influence decisions and funding throughout the American educational infrastructure. And, as Dunbar implies, it is going to be a prodigious task to define "citizenship" and to define and then measure "global competitiveness," leave aside trying to relate (in an academically rigorous fashion) the achievement of values that underlie these realms to performance on a test.

Nonetheless, *these are the premises of Goal 5*, and thus they formed the nucleus for discussion of the national assessment of college student learning. One important threshold issue raised by Paul and Nosich is the value of critical thinking skills that generalize, that apply to the widest possible frame of reference. Not merely academic hairsplitting they suggest, certain generalizable skills provide a fundamental way of conceptualizing a successful civilization, and ergo the educational system that prepares its members. They urge the widest and most socially relevant frame in defining the higher skills, citing David Kennedy, the president of Stanford University, in a speech to 3000 college and university presidents:

> It simply will not do for our schools to produce a small elite to power our scientific establishment and a larger cadre of workers with basic skills to do routine work. Millions of people around the world now have these same basic skills and are willing to work twice as long for as little as one-tenth our basic wages. To maintain and enhance our quality of life, we must develop a leading-edge economy based on workers who can think for a living. If skills are equal, in the long run wages will be too. This means we have to educate a vast mass of people capable of thinking critically, creatively, and imaginatively.[348]

Paul and Nosich are careful to avoid the sterility of the excessively academic viewpoint, defining critical thinking skills as those that are "seen as valuable by practitioners of the academic disciplines, by responsible leaders of government, of the professions, of business, by citizens interested in their environmental, physical and economic welfare."[349] Core abilities which manifest themselves in such value-laden ways, they say, are "ways to adapt to rapidly changing knowledge, to recognize problems and see their implications before they become acute, to formulate approaches to their solution that recognize legitimately different points of view, to draw reasonable conclusions about what to do."[350]

---

[348] Paul and Nosich, op. cit., 8.

Be wary of too many abstractions, they warn. "The assessment should contain items that, as much as possible, are examples of the real-life problems and issues that people will have to think out and act upon."[351] Nummedal played this theme repeatedly and with some force. "It may be helpful to remember that when specific courses in critical thinking have been introduced into the curriculum, one of their major goals has been to enable students to think more effectively about everyday issues and concerns, such as practical problem solving and decision making. Indeed, some have even argued that this is *the* most important goal of such a course."[352]

While Nummedal suggests the entire national assessment of college student learning process be re-framed with this more "practical" orientation, Paul and Nosich have more confidence in the current state of critical thinking, at least its rich conceptualization. They prefer to focus on the Goal 5 aspect of preparation for citizenship, and urge that national assessment of college student learning standards be framed that way:

> Both public and private life increasingly require mastery of the basic skills and abilities of critical thinking. When this mastery is absent, the public degenerates into a mass society susceptible to manipulation by public relations specialists who can engineer political victories by an adroit use of mud slinging, scare tactics, shallow nationalism, fear, envy, stereotype, greed, false idealism, and maudlin sentimentalism. Modern citizenship requires basic critical thinking skills and abilities throughout.

---

[349] Ibid., 8.

[350] Ibid., 8.

[351] Ibid., 4.

[352] Nummedal, op. cit., 4. For examples, she points to R.J. Sternberg, "Questions and answers about the nature and teaching of thinking skills," In eds. J. B. Baron and R. J. Sternberg, *Teaching thinking skills: Theory and practice*, (New York: Freeman and Co., 1987): 251-259.

> The modern citizen should be able to assess the arguments presented for his or her assent, must rationally adjudicate between conflicting points of view, must attempt to understand a culturally complex world, must assess the credibility of diverse sources of information, must translate between conflicting points of view and diverse appeals, must rationally decide between priorities, must seek to understand complex issues that involve multiple domains (for example, the environmental, moral, economic, political, scientific, social, and historical domains). Without a solid grounding in the skills and abilities of critical thinking, citizens are intellectually disarmed, incapable of discharging their civic responsibilities or rationally exercising their rights.[353]

The generally lamentable state of American students *vis-a-vis* critical thinking reflects a lack of emphasis directed toward such skills. National assessments, say Paul and Nosich,

> in virtually every subject indicate that, although our students can perform basic skills pretty well, they are not doing well on thinking and reasoning. American students can compute, but they cannot reason. They can write complete and correct sentences, but they cannot prepare arguments. Moreover, in international comparisons, American students are falling behind—particularly in those areas that require higher-order thinking. Our students are not doing well at thinking, reasoning, analyzing, predicting, estimating, or problem solving.[354]

As Morante, Lenth, and others familiar with the college assessment experience pointed out, observing such a distinction between basic and higher skills is one thing, implementing it in a democratic system predicated on great access to American higher education is quite another. Ratcliff points out the connection between this fact of American life, and the "competitiveness" aspect of Goal 5:

> A frequent rationale for a collegiate-level testing program is that it will help insure our global competitiveness by having better educated workers. It should be noted that in other industrialized countries with whom we compete, including Germany, France, Great Britain and Japan, national testing programs work to exclude all but a small

---

[353] Ibid., 7.

[354] Paul and Nosich, op. cit., 1.

proportion of students from going to college. College-going rates in these countries range from 17 to 25 percent.[355]

Thus there could be a problem of comparing apples and oranges, even with respect to other democracies. France and Great Britain, says Ratcliff, are working to increase access to higher education and have set a goal to double the number of those advancing. One must pose the unanswerable and paradoxical question, in light of Goal 5. Are certain of these countries more "globally competitive" than America because of our egalitarian principles, as manifested in greater access to higher education? Or, as Ratcliff puts it,

> It is no small irony that our competitors in the industrialized world are seeking to create more open and accessible higher education systems at a time when we seek to contract, exclude and be more selective. Surely global competitiveness cannot be further[ed] by both greater selectiveness and greater access. Are we on the right path to global competitiveness? We desperately need to examine what we mean by that term before accepting it blindly as a rationale for educational reform.[356]

Others pointed out the dangers of a facile premise for overhauling the American system. Miller reminds us that "the national goals were developed in part because of a wide-spread perception that American workers are becoming increasingly less competitive in the world economy."[357] Thus looking at "crude measures of job success in that American market" as it is now operating may not be particularly predictive about how—to ensure global competitiveness in the future [or, more accurately, comparably advanced workplaces]—we should go about changing the educational preparation of workers in the future.

Notwithstanding this important caveat, the good news from the critical thinking movement is that "critical thinking skills and abilities are highly transferable to the workplace."[358] Not only as a basic constituent of the underlying intellectual standards (earlier referred to) that would apply to any context, but because the workplace can be viewed as the ideal paradigm context for critical thinking—where goals are sought through the solution of problems. "Moreover," Paul and Nosich say,

---

[355] Ibid., 16.

[356] Ratcliff, op. cit., 16.

[357] Miller on Morante.

[358] Paul and Nosich, op. cit., 6.

> the kind of "work" increasingly required in industry and business is "intellectual," i.e., requiring that workers define goals and purposes clearly, seek out and organize relevant data, conceptualize those data, reason to legitimate conclusions, consider alternative perspectives, adjust thinking to context, question assumptions and modify thinking in the light of the continual influx of new information.[359]

As mentioned before, such analyses from Paul and Nosich can be referred to the discussion on identifying skills, but it seems more fruitful to look at certain of their criteria for effective critical thinking as a question of standards. For instance:

> Furthermore, the intellectual work required must increasingly be coordinated with, and must profit from the critique of fellow workers. There is no avoiding the need, therefore, to express ideas well, accurately represent and consider fairly the ideas of others, write clear and precise memos and documents, and coordinate and sequence all of these so that well-reasoned policies and decisions can be accurately understood and effectively implemented.[360]

Because Cappelli's research focus is on the workplace, he understandably worries about how effectively whatever standards that may be developed for a national assessment of college student learning will reflect the real values he believes are evident there. He thus reinforces the need to conceptualize the critical thinking assessment in a way that fundamentally differs from the traditional ways that students have achieved in the past. It is easy to see, he explains, why "grades may not be good predictors of job performance, even for subjects where the course material may be relevant to jobs." The crucial factor is *what skills* are relevant to job performance. Courses, he believes, generally "do not teach skills relevant to jobs and. . .grades are not based on those skills even where they are taught."[361]

---

[359] Ibid., 6.

[360] Ibid., 6.

[361] Cappelli, op. cit., 9.

In particular, certain structural aspects of higher education serve to confound the translation of standards to the workplace: large lecture formats with little conversation and discussion among students or with professors; emphasis on memorizing results of prior research, and unimaginative presentation of such course material; and multiple-choice tests. While these elements of postsecondary education may seem to be functional or structural, Cappelli believes their impact goes to the heart of the quality and suitability of graduates for the workplace. He paints an alternative vision of a college course, for example, in human behavior:

> Consider the [course being taught] in a small group discussion format where students do at least some of their work in teams; where the material requires students to apply theories and statistical methods to real life problems; where grades are based on written efforts to evaluate critically course material and on class participation. [In this case], the education process develops many useful skills, and the grading procedure can evaluate them.[362]

Cappelli's emphasis on flexibility dovetails with his view of how the entire national assessment of college student learning process (given the years of developmental time before it actually begins to enhance the workplace) must build into its very structure the standards and values, the skills and abilities, that are predicted to transfer to the workplace of the future. His paper analyzes several types of data about *changes* in the nature of *future* jobs. While he concurs with a number of writers who argue "that the distribution of jobs in the economy is shifting away from low-skill positions such as manual work and toward higher skill jobs like engineering," he reads the consensus about these studies to indicate "that while there is likely to be a shift in this direction, the rate of change will be no greater than in past generations."[363]

Some of his research, however, is particularly relevant to the national assessment of college student learning, as it focuses not so much on broad job categories as on how jobs might be changing, how their underlying knowledge, skills and abilities may be evolving to reflect different standards and values, "how management jobs may be different in the future."[364] And while such analyses are often difficult to mount objectively, "certainly there is a consensus that managerial jobs have become less secure and that the ranks of managerial jobs have been thinned, leaving more work for those who remain."[365]

---

[362] Ibid., 8.

[363] Ibid., 9.

Nonetheless, work has been done in a series of interviews to establish some credible predictions about the winds of change in the workplace. Cappelli cites a 1988 study by Porter and McKibbin conducted for the American Assembly of Collegiate Schools of Business "that considered how businesses were changing and the implications for jobs and education oriented toward management jobs. Their conclusions from extensive interviews suggest that education needs to be more applied—help students see the links to practice—and that interpersonal and leadership skills should be emphasized [which are] oriented toward managing people."[366] A similar study for the SEI Center at Wharton sought to develop recommendations for revising their curriculum toward this future-oriented view. "The recommendations included more extensive training in interpersonal skills, greater integration across disciplines, and more breadth in education."[367] Cappelli and his colleagues

> also conducted interviews with human resource consultants in firms that specialize in job analyses to get their thoughts on the future requirements of jobs. There is a clear consensus that flatter organizations with less hierarchy are forcing employees to be more autonomous. The reduction in structure and control associated with it implies greater reliance on leadership skills as the alternative for managing employees. Communication skills are also becoming more important as employees have more informal reporting arrangements with more people and as matrix organizational structures and team methods of work organization force employees to work more with each other. Interpersonal skills in general become more important as working in teams becomes more prevalent. The ability to be flexible and adapt to new

---

[364] Ibid., 9; see Peter Cappelli, "Assessing Changes in Skill Levels for Production and Clerical Workers, 1978 - 1988," (Philadelphia: National Center on the Educational Quality of the Workforce, 1991.)

[365] Ibid., 10; see Peter Cappelli, "Examining Managerial Displacement," in *Academy of Management Journal*, (In press).

[366] Ibid., 10.

[367] Ibid., 10; see Jerry Wind and Alfred P. West, Jr. "Reinventing the Corporation." *Chief Executive Magazine,(In press).*

circumstances is another general theme that is driven by the continuing turbulence in modern corporations.[368]

Thus does the question of standards begin to come a bit more into focus. In lieu of abstract ideals that might require tortuous intellectual translation into a practical context, many national assessment of college student learning advisors logically looked at the question of transfer from the "outside in:" Work toward a national assessment of college student learning to be structured around standards that match those implicit in the goals (that is, globally competitive, enhanced critical thinking skills, and better citizens). Whether an instrument that did embody such standards would really indicate that the abilities being demonstrated would themselves transfer is another prickly question, and takes the debate into the realm of content domains, and subject-specific instruction.

d. More specifically, must a national assessment of college student learning test subject-specific content domains in order to generate robust and reliable conclusions about transfer?

"Target skills must be measured *in context,*" insists Dunbar: "Any context-free surrogate developed in the interest of efficiency and expediency will not only fail to satisfy the audience for a national assessment of college student learning, but will also suffer from major shortcomings with respect to validity, broadly defined."[369] This conclusion seemed to reflect a majority view among the workshop participants. Without the complexity and subtlety afforded by meaningful subject content, most believed, the Objective 5 abilities can't be probed. As Dunbar puts it: "generalized skills of this nature are not meaningfully taught, nor measured, *in a vacuum*. Perhaps by definition, some degree of complexity *in content* is needed in order to exercise a desired level of complexity in the cognitive *processes* of critical thinking, communication, and problem solving."[370] This point isn't merely impressionistic, but was revisited over and again by participants concerned about transfer. The Alverno model goes to great lengths to validate the transfer of their skills into the subsequent work setting, and Dunbar himself cites Larkin who "traced developments in cognitive psychology on the question of transfer and concluded that

---

[368] Ibid., 10.

[369] Dunbar, op. cit., 8.

[370] Ibid., 3.

> Although attractive, the notion that transferable knowledge is a core of general problem solving skills has been historically unproductive. There is not good evidence that instruction in such skills improves performance. . .There is evidence of varying kinds and of varying strengths that skills that are somewhat domain specific may transfer. These include strategies that apply to a moderately broad range of domains, skills for managing the surroundings of a task, and skills for learning. None of these kinds of knowledge, however, forms a complete routine that can be executed in the absence of other knowledge; all are intermingled with other domain-specific knowledge.[371]

Dunbar notes that there is

> a body of research in cognitive psychology that portends difficulty with respect to validity when conducting general—rather than subject-specific—assessments of higher order skills and expecting these results to transfer to particular applications at a later point in time. If these writers are correct, then from a measurement point of view, it is unlikely that any instrument [measuring critical thinking, communicating effectively, and problem solving as general intellectual skills will] satisfy accepted standards for the content and construct validity of educational measurements, particularly those that are used for external assessments of educational progress.[372]

Dunbar admits that while "it is probably misleading to say that [Larkin's] point of view is universally accepted as a matter of fact among cognitive psychologists, the recognition of uncertainty with regard to transfer of higher-order skills is characteristic of writers in the field: "much important cognitive activity is domain specific. . .it seems likely that even general problem solving strategies are conditioned by or adapted to the particular characteristics of the knowledge domain in which they are used."[373] "If these writers are correct," Dunbar continues, "then from a measurement point of view. . .any instrument that proposes to measure critical thinking, communication, and problem solving as general intellectual skills is not likely to satisfy accepted standards for the content and construct validity of educational measurements, particularly those that are used for external assessments of educational progress."[374]

---

[371] Ibid., 14. Dunbar cites J.H. Larkin, J. H. "What Kind of Knowledge Transfers?" in ed. L.B. Resnick, *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, (Hillsdale, NJ: Erlbaum, 1989): 383-4.

[372] Ibid., 14.

[373] Dunbar citing R.E. Snow and D.L. Lohman, "Implications of cognitive psychology for educational measurement," in ed. R.L. Line *Educational Measurement* (New York: Macmillan, 3rd. ed., 1989): 305.

And an even greater uncertainty is whether general as opposed to subject-based skills can be captured in a measurement that is content valid. "General education, as a curriculum in higher education and as a construct for educational measurement, simply lacks the focus that is truly necessary for the development of instruments that are content valid in the judgment of experts."[375] And those measures that have been developed to measure "the outcomes of general education (e.g. ACT-COMP, CAAP, and Academic Profiles) have been the subject of much criticism in the literature on assessment in higher education as much because of the ill-defined nature of the domain as because of internal limitations of the measures themselves."[376]

Nonetheless, emphasizes Nummedal, this assumption is at the heart of the national assessment of college student learning process, and must be examined:

> [It] has been described variously as the goal of "transfer," "generalization," or "application." Concern with this goal has led, for example, to the controversy over the relative benefits of teaching general reasoning processes vs. domain specific processes.[377] It is inherent in questions such as: "How can logic be made more relevant to everyday reasoning?"[378] and "What can one do to maximize transfer of training for thinking skills to students' everyday lives?"[379]

Mayer's 1991 survey, *Thinking, Problem Solving and Cognition*, reinforces this view of *general* intellectual skills in accounts of creativity training and expert problem solving:

---

[374] Dunbar, op. cit., 13.

[375] Ibid., 12. Experts such as D.B. Yarborough, "*Assessing cognitive general education outcomes: Conclusions from a decade of research on the ACT COMP measures*," Appendix E: Supplement to the COMP Technical Report, (Iowa City, IA: American College Testing Program, 1991).

[376] Ibid., 12.

[377] R. Glaser, "Education and thinking: The role of knowledge," *American Psychologist, 39(1984):* 93-104.

[378] J. Baron, *Thinking and deciding.* (Cambridge: Cambridge University Press, 1988): 153.

[379] Nummedal, op. cit., 4; citing R.J. Sternberg, "Questions and answers about the nature and teaching of thinking skills," in eds. J.B. Baron and R.J. Sternberg *Teaching thinking skills: Theory and practice* (New York: Freeman and Co., 1987): 251-259.

> Despite the claim of the classical approach that general creativity can be taught, most objective studies show that students learn component skills that can be used mainly on problems like those given during instruction. At present, after sixty years of experience with creativity-training courses, there is not convincing evidence that global skills can be learned in context-free environments. In short, there is no quick way to improve general problem solving performance.[380]

The study of expert problem solving further underscores the point that knowledge-based thinking can be distinguished in almost every way from so-called general intellectual skills. Why are experts so much better at thinking, in their fields of expertise, than novices? While the answer may seem tautological, the significance for the national assessment of college student learning should probably not be underestimated. "The problem-solving difficulties of novices can be attributed to inadequacies of their knowledge bases and *not to limitations in the architecture of their cognitive systems or processing capabilities.*"[381] (Emphasis added) Weinstein speaks from a major center of critical thinking research at Montclair State:

> It's clear that problem solving requires critical thinking. It's clear that critical thinkers ought to be adept at problem solving. . .It's also clear to people who know the tradition that problem solving has been developed within a discourse community that is far different from the discourse community within which critical thinking has been developed, and developed through engineers who have very different senses of how problem solving ought to be articulated, how it ought to be manifest, and how it ought to be measured.
>
> For example, they use mechanical and technical problems, and don't use issues of political and cultural concern. Similarly, the generic skills represent truly universal areas of concern that all thoughtful people should be able to address in responsible ways. But how these areas of concern and how these skills are articulated, manifested, and assessed might look very, very different from the points of view of

---

[380] Richard E. Mayer, *Thinking, Problem Solving and Cognition*, (New York: W.H. Freeman and Company, 2d ed., 1991): 386.

[381] M.T.H. Chi, R. Glaser and E. Rees, "Expertise in problem solving," in R.J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence*, Vol. 1 (Hillsdale, NJ: Erlbaum, 1982): 71.

> people who work in discourse frames as diverse as the physical sciences and the humanities.[382]

Venezky also reviewed part of the literature from this cognitive perspective, and reaches the same conclusion in his survey of the assessment of literacy. Asking "whether or not higher order literacy can be assessed independently of a particular content area," he cites recent work in curricular assessment and reading comprehension to suggest that higher level literacy skills probably cannot.

> After wandering through the presentation on assessment options, I am convinced that lower level reading skills can be evaluated independently of specific curricular areas but that higher level literacy skills probably cannot. At the highest levels of literacy assessment are those skills that require integration of text-derived information with information obtained previously or from other texts in the same task. Although artificial situations might be created, drawing upon "neutral" content, I doubt that meaningful items can be constructed by this means. Therefore, to eliminate some of the confounding of content area competence with literacy ability, I suggest that some of the higher level tasks provide content area background materials along with the test passages, and that these be located primarily in each student's major area.[383]

Dunbar forces home the point by concluding that "a meaningful assessment, one that truly responds to Goal 5, must gather information about higher-order skills within the context of specific disciplines. Of interest is not the content of the chemistry or sociology curricula, but rather instances in which evidence of critical thinking, communication, and problem solving is transparent from that content."[384] However, points out Nummedal, since explicit critical thinking instruction is not widespread in postsecondary education, the national assessment of college student learning for the near-term must in some sense bow to current realities.

Nummedal does not believe that, if critical thinking instruction were more advanced and widespread, it would obviate the debate:

---

[382] Mark Weinstein in Open Session.

[383] Venezky, op. cit., 30.

[384] Dunbar, op. cit., 18.

> The notion of explicit instruction in critical thinking has been the subject of continuing debate. Much of this debate is over the issue of whether critical thinking is discipline specific or can be taught generically. This debate is far from over. . .If we look at the higher order thinking skills that are being taught currently in college, my sense is that, with some notable exceptions (e.g., Alverno College), we are looking at discipline specific conceptualizations of critical thinking and problem solving. Even on campuses where specific coursework in critical thinking has been mandated (e.g., campuses in the California State University System), courses meeting the requirement are taught in different departments (most often philosophy and psychology). These departments have courses which derive from specific frameworks of critical thinking tied to theoretical formulations grounded in a specific disciplines.[385]

Facione counters that, in fact, the debate is over:

> It is the general consensus among critical thinking theorists and assessment experts that this is a dead issue. Robert H. Ennis, perhaps the leading figure in critical thinking and critical thinking assessment in the nation, handily laid that matter to rest some years back. There may be some nicely refined twists put on some critical thinking applications in advanced doctoral research done within given disciplines, but for the college level the list of core critical thinking cognitive skills (analysis, interpretation, evaluation, inference, explanation, and self-regulation) are both practical and generic. The critical thinking dispositions are as well.[386]

Boehm interprets Paul and Nosich—and endorses the concept—as commending national assessment of college student learning developers to include items addressing both models. "However, if the idea fails to win support, then I would argue for a test that assesses critical thinking defined as understanding and applying at the appropriate gradient the modes of inquiry, the language, the thinking, done by practitioners of the disciplines." Critical thinking "is best understood as doing the mental work of the disciplines, and is best taught in all courses across the curriculum by teachers who have 'unpacked' the thinking required by their discipline, organized it gradiently, and have made it the backbone of their courses. However, most critical thinking teachers now agree that doing both is optimal."[387]

---

[385] Nummedal, op. cit., 10.

[386] Facione on Nummedal. "Professor Ennis directs the Illinois Critical Thinking Project out of the Department of Educational Policy at the University of Illinois at Urbana-Champaign. That unit has been conducting research on both everyday content and discipline-specific contexts for CT assessment for a number of years."

Michael Scriven saw this same duality "emerging from the [workshop] discussion which I want to utter a caution about—

> The format is: there should be the general critical thinking section of the master test, whatever it is, and then there will be the subject matter-specific sections. And this is our kind of bowing in the direction of the importance of critical thinking in the disciplines, which is indeed an important matter. . .That is, I think that the contribution of good critical thinking instruction to problem solving in particular but to critical thinking in general, is mastering the general methodology of half a dozen general areas.[388]

He would operationalize this goal by providing test takers "an option of your choice of interpreting poetry, or doing analysis of thermodynamic phenomena" but then indicate "massive extra credit is available if you can do them all. . .

> I think it's important to [consider] a reasonable part of problem solving and critical thinking [to] involve mastering a large number of methodologies. . .I think we all ought to be literate with respect to the notion of social science control groups, to the notion of lab standards, and measurement procedures, and observable errors, and so on and so on. It's not that hard, but it's something we won't all master at the age of sixteen. But it's something we ought to keep working towards.[389]

The discipline-specific issue nonetheless poses something of a juggling act for national assessment of college student learning developers, say Resnick and Peterson. "The majority of colleges and universities, for historical reasons, educate their students through majors in specific disciplines. What a college student should know and be able to do depends largely upon his or her field of study. The search for students more capable of critical thinking will thus depend in some way on the quality of study in the major field."[390] But the richer critical thinking model requires more, and most discussants concurred with Resnick and Peterson that "knowledge of fields outside the major field is also essential. The college graduate should have a deep understanding of his or her major area of study, but should also be able to relate knowledge to other disciplines. This view is well expressed in the Association of American Colleges publication, *The Challenge of Connected Learning:*

---

[387] Boehm on Paul and Nosich. "This is decidedly the case at my own institution [Oakton Community College], where for six years faculty worked successfully to build the "*intra*disciplinary" model, and where recently they have begun to develop a course in the "*inter*disciplinary model, which is now seen by them and just about everyone else as a compliment to what came before and continues."

[388] Michael Scriven in Open Session.

[389] Ibid.

> To fulfill its role in liberal learning, the major must also structure conversations with the other cultures represented in the academy, conversations that more nearly reflect the diversities within our world and require patient labors of translation.[391]

"The problem of translation across fields has grown with knowledge in the disciplines," say Resnick and Peterson. "Each student needs experience in both interpreting the meaning of work in other fields and conveying the meaning of his or her own field to others. The kind of skills and integration that we assumed in goal five depend on that training and capability."[392]

The recognition that to achieve valid transfer one must test specific content domains, many felt, would refocus attention on how specific majors and college programs would, could and should relate to any national assessment of college student learning. Resnick and Peterson point out that "any effort to raise the exit level skills of these students must recognize that such know-how is subject matter dependent and needs to be cultivated through a challenging curriculum that shapes the learning of our undergraduates and their major programs."[393] Ratcliff's work at the National Center for Postsecondary Teaching, Learning and Assessment was discussed in the first chapter as a possible program to improve and inform instruction.

Ratcliff cites other such efforts, such as that at "the Center for Assessment Research and Assessment at the University of Tennessee [that are] developing standards of good practice that could also be applied to classroom assessment practices as well. Institutions could be encouraged to develop these indirect indicators of student learning in the three key areas of assessment. These indicators would not only provide initial estimates of how students are progressing as well as the extent to which current college curriculum is directed to the enhancement of critical thinking, clear communications and problem solving abilities."[394] These and similar programs, he believes, could provide a more direct database for national assessment of college student learning development, what he calls "an institutional profile which, while useful for discerning national patterns and trends, could always be used as a final, sensitive filter to accommodate the uniqueness and idiosyncracies of particular contexts."[395]

---

[390] Resnick and Peterson, op. cit., 7. "Our proposal for goal five indicators recognizes this fact."

[391] American Association of Colleges, *The Challenge of Connected Learning* (Washington D.C.: American Association of Colleges, 1991): 5.

[392] Resnick and Peterson, op. cit., 7.

[393] Ibid., 7.

Thus, while much research and a majority of discussants at the workshop lined up behind the premise that content domains must be subject-specific, Lenth raises a dilemma. His familiarity with the university infrastructure in America suggests that content domains currently provide the focus for most state and institutional evaluation programs. Lenth throughout serves the national assessment of college student learning process by casting a wary eye on how any proposals that are developed might actually be received in the postsecondary educational community. Since the point was made repeatedly throughout the papers and discussions that faculty must get behind any national assessment of college student learning for it to be successful, he worries that such "interest groups" may feel their hegemony threatened by a serious probe into assessment which relies on content. "Ongoing state and institution assessment activities focus primarily on the teaching and learning environment, and are weakest in the areas of general outcomes measures."[396]

Morante earlier described some of the disruptions in the assessment process experienced in the New Jersey GIS test in 1990, even though in that assessment the SLO Committee had anticipated Lenth's recommendation by building its instrument around general outcomes measures. First, "no attempt was made to relate the content of the tasks to the student's major or to courses completed. The emphasis is on assessing the underlying general intellectual skills needed by all students regardless of major or institution."[397] This does not refute those who contend that such general skills don't transfer, nor does it say that critical thinking as called for by Goal 5 or a national assessment of college student learning was effectively measured by the New Jersey GIS. It *does* say, however, that statewide administration of an instrument with criterion-referenced measurement of [primarily] tasks instead of multiple choice responses can be accomplished. ETS evaluated the development of the test, concluding:

---

[394] Ratcliff, op. cit., 25.

[395] Ibid., 26.

[396] Lenth, op. cit., i.

[397] Morante, op. cit., 17.

- The materials developed for the assessment of general intellectual skills, especially the extended tasks, are valid and innovative measures of certain college level intellectual skills.

- The extended tasks and scoring processes "worked," that is, students could respond and readers could score them reliably. The GIS Assessment is therefore, an appropriate measure of the skills its seeks to assess.[398]

So the more general the measure and the more categorical its treatment, argues Lenth, the better it fits the current American testing scene. Give the colleges what they do not have or are not now doing well, he predicts (and his survey shows this to be: cognitive testing of higher order skills, reported with general outcome measures) and they will be less likely to feel it as a threat, and more likely to provide the crucial support at the interface with students that all agree is essential for success.

The question is not likely to go away. A strong majority, backed by a wealth of data and studies about transfer, believe that for an assessment to have any pretense of validity it must be discipline-based. However, political and institutional attitudes and environments make this problematic. Thus, if a national assessment of college student learning evolves as strongly discipline-based, Lenth's concerns about how it will impact faculty give rise to a number of practical issues, which in turn suggest a number of policy questions, perhaps the largest of which is accountability.

In Chapter 1, a number of authors made clear that their analysis was predicated on a political view of the consequences of a national assessment of college student learning. Put simply, if the test shows that students can't demonstrate what it is that those setting standards think they should know, what will happen? This question is only partly rhetorical. Obviously, fingers will be pointed, and blame assigned. At a more substantive level, courses, curricula, and content will be re-examined for their contribution to the failure, and their potential for remediation. Ratcliff, Cappelli, and others see this as an opportunity; Lenth, personally acquainted with many of the people who will feel the impact of the first shells, warns that it could doom political acceptance of the entire undertaking if proper collaboration isn't secured and the groundwork laid first. The many issues involved have not yet been systematically considered, but Morante's in-depth history of the New Jersey experience provides insight into the nature of the problem.

---

[398] Educational Testing Services, *A Report on the Development of the General Intellectual Skills Assessment*, (Princeton, NJ: Educational Testing Services, 1989): 2.

e. Is a single set of standards reasonable, or even possible, given the diversity of institutions and of those to be assessed?

Validity in this more scientific sense is not limited to questions of transfer. Banta framed her analysis around an attack on the premises underlying a national assessment of college student learning, one of which is that "reliable and valid measures of student achievement of the defined abilities can be defined or created. . .measuring what they purport to measure."[399] Her experience at the University of Tennessee, Knoxville (UTK) and the Center for Assessment Research and Development suggests otherwise. "Over the past thirty years, measurement theorists have spent considerable amounts of time and energy debating the issue of whether skill in critical thinking is more dependent upon deep expertise in a specialized area or upon possession of well-developed generic reasoning strategies."[400]

And Banta's group at UTK has been pilot testing multiple instruments with seniors to probe various of their technical qualities and ultimate reliability, especially with respect to that distinction.[401] Only one in four UTK seniors themselves believed the test(s) to be a "good or excellent measure of their knowledge and skills in general education, and faculty concluded that none of the tests assessed more than one-third of the content specified for inclusion in the University's general education program."[402] Banta's conclusions anticipate the next question in this chapter: whether anything like a standard instrument can reveal the micro abilities Paul and Nosich insist comprise the tools of critical thinking; whether they can penetrate to skills that have been learned in college as distinguished from abilities students brought with them:

---

[399] Banta, op. cit., 1.

[400] Ibid., 11; see D.N. Perkins, and G. Salomon, "Teaching for transfer," in *Educational Leadership* 46(1)(1988): 22-32.

[401] Ibid., 10. "All four of the standardized exams under scrutiny purport to measure some dimensions of the critical thinking and communication skills of college students. Scales on the COMP exam include Functioning in Social Institutions, Using Science and Technology, Using the Arts, Communicating, Solving Problems, and Clarifying Values. The CAAP includes Writing, Mathematics, Reading, and Critical Thinking scales. The Academic Profile is composed of scales labeled College-Level Reading, College-Level Writing, Critical Thinking Using Mathematical Data, Humanities, Social Sciences, and Natural Sciences. Scales that make up the CBASE are English, Mathematics, Social Studies, Science, and interpretive, strategic and adaptive reasoning."

> Regardless of the measurement approach utilized, however, our studies show that students' scores on all four tests are much more highly related to initial ability than to any other factor. Attempts to trace the impact on these scores of coursework and other educational experiences associated with the college years have not yielded definitive answers.[403]
>
> Hanson[404] attributes this failure to the fact that today's test developers know best how to measure static traits, such as verbal ability, as opposed to developmental changes. Since measures of static traits are based on the assumption that the underlying structure of the construct being measured does not change over time, such measures may not be able to detect student characteristics that change as a result of college experiences.
>
> The evidence assembled to date from research and experience in postsecondary outcomes assessment leads to the conclusion that current measurement theory and its application in the development of instruments designed to assess students' general intellectual skills are inadequate to support specific suggestions for improving students' learning based on their scores on these instruments.[405]

If initial, general ability is what—finally—is being tested for, then Banta's assault on approaching the national assessment of college student learning as a newly-developed, one-time instrument must be given more serious consideration. Because such abilities (the American educational establishment has come to accept) generally reveal more about the educational and cultural background of the test taker than they do about critical thinking abilities in their richest sense. So while validity in an assessment thus would seem to require a focus on specific content domains, thus far it has only been suggested that this is a necessary, not a sufficient, condition for a valid test. Objectivity, however, is a horse of a different color, which raises the enormous specter of society's relation both to its professed aims as specified in Goal 5, and to imposing judgements and conclusions about the achievement of such goals onto a culturally and educationally diverse population.

---

[402] Ibid., 11.

[403] Banta here cites G.R. Pike, "Using mixed-effect structural equation models to study student growth and development," Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, 1991.

[404] G.R. Hanson, "Critical issues in the assessment of value added in education," In ed. T. W. Banta *Implementing outcomes assessment: Promise and perils. New Directions for Institutional Research, 59* (San Francisco: Jossey-Bass, 1988).

[405] Banta, op. cit., 11.

Imposing Social Judgments about Value

The declaration of value, emphasizes Dunbar, is what is revealed behind the curtain, and national assessment of college student learning developers must confront this dilemma head-on:

> There is no value-free and thereby objective scale for the kind of measurement that *America 2000* envisions for higher education in the United States. Moreover, any scale that is developed for national assessment of college student learning will ultimately be interpreted with respect to some such scale of social utility, because that is the real foundation of the goal statement from *America* 2000, whether we like it or not."[406]

Dunbar, in addition to Rock and others, invoked the recent modifications made to the NAEP mathematics assessment as a harbinger of this sort of dilemma. "The recent controversy surrounding the attempt by the National Assessment Governing Board (NAGB) to set this kind of standard for a comparatively well-defined domain in the NAEP Mathematics Assessment[407] might well pale in comparison to the disagreements likely to arise in determining whether or not gains in, say, critical thinking skills are substantial enough."[408] NAEP, reflects Dunbar, began with a very modest purpose, to provide descriptive information to the public concerning educational progress, and it responded to the charge with simple reports that emphasized concrete examples of exercises and results, much like a public opinion poll would do. Like the Gross National product or the Consumer Price Index, NAEP provided the data, but wasn't itself a policy instrument.[409]

---

[406] Dunbar, op. cit., 5.

[407] Dunbar cites R. Rothman, "NAEP board fires researchers critical of standards process," in *Education Week*, (11)1(1991): 36.

[408] Dunbar, op. cit., 5.

Hambleton in his review recounts some of the issues Dunbar referred to, suggesting that "many principles were learned about the selection and training of judges, data analysis, and other aspects of the standard-setting process that could be useful in future national standard-setting initiatives." However, he concludes, "the NAGB effort was expensive, time-consuming, and many measurement problems arose in the process," problems [he believes] that would pale compared to standard-setting "in the area of critical thinking with its complex definitions, ambiguities, performance assessments, and arbitrary scoring methods."[410] But [perhaps inevitably] NAEP's success as an indicator prompted policymakers at NAGB to "experiment with the added political responsibility of judging the adequacy of achievement by setting standards that in effect attempt to map the achievement scale onto the utility scale. National assessment of college student learning," continues Dunbar, because it is conceived "as an endeavor that responds directly to the *America 2000* goal for college students, should be recognized from the outset as having been charged with a judgmental responsibility."[411]

The next logical question becomes, What are you trying to judge? More specifically, What will the standards (both the inherent standards of critical thinking, and the achievement standards imposed on the assessment) reveal about the relative "qualities" of those being tested? Is there a single answer to such a question?

The "Multiple Measures" Dilemma

Ratcliff would say no. "There are multiple visions of what constitutes intelligence and learned abilities. Enlightened approaches to assessment include multiple definitions of learned abilities and multiple measures of that learning."[412] At the very least, such a conclusion builds on the insights gathered under content domain testing, and suggests that the national assessment of college student learning "avoid the essentialistic quest for the one best set of measures that [would] encompass all of general learning and cognitive development at the collegiate level. . .So long as we believe that studying different subjects produces different types of learning, and so long as higher education forwards a curriculum that attempts to embody the expanding diversity and complexity of advanced human thought, multiple definitions of the ingredients to intelligent behavior and multiple means to assess them will be required."[413]

---

[409] Ibid., 6.

[410] Hambleton on Dunbar.

[411] Dunbar, op. cit., 6.

This broader, richer, mosaic view of the target was also taken by Nummedal, whose vision of "collegiate level" achievement cannot necessarily be reconciled with Goal 5:

> We should be aware that the kinds of skills and dispositions necessary to meet the goals of success in the workforce, effective exercise of the rights and responsibilities of citizenship, and life-long learning most likely will not be well developed in the typical college graduate of 22 years of age, not even in the 30 year old graduate. Attributes associated with reflective judgment (i.e., recognition of the limitations of personal knowledge, acknowledgment of the general uncertainty that characterizes human knowing, and humility about one's own judgments in the fact of such limitations) involves an epistemological stance rarely found in young college graduates.[414]

---

[412] Ratcliff, op. cit., 14, citing R.J. Sternberg, "The tyranny of testing," *Learning*, 17(March, 1989): 60-63; and work done with his colleagues:

J.L. Ratcliff, "Development and testing of a cluster-analytic model for identifying coursework patterns associated with general learned abilities of college students: Final report, May 1990 U.S. Department of Education, Office of Educational Research and Improvement, Research Division. Contract No. OERI-R-86-0016. University Park, PA: Center for the Study of Higher Education, Pennsylvania State University.

[413] Ibid., 14.

[414] Nummedal, op. cit., 16. K.S. Kitchener and H.G. Brenner, Wisdom and reflective judgment: Knowing in the fact of uncertainty, In ed. R.J. Sternberg, *Wisdom: Its nature, origins, and development* (Cambridge: Cambridge University Press, 1990).

A third author talking about multiple measures was Dunbar, who anticipated a major theme of the workshop: the implications for testing of the diversity of the universe being tested.

> With a system of postsecondary education characterized by pluralism and open access, though with varying degrees of selectivity and admissions standards, a decision must be made regarding the validity of a single achievement standard for all types of institutions. Because colleges and universities in the United States have different sources of funding and see their particular missions in higher education from sometimes unique perspectives, perspectives known to their applicants, it may be that standards will need to be conditioned on the mission statements of colleges and universities.[415]

The implications for the national assessment of college student learning, believes Dunbar, are that "NCES may need to develop multiple standards for multiple institutions, particularly if reports of results are disaggregated to the institution, which is the direction NAEP has moved in recent years."[416]

This further reference to the NAEP experience suggests that policymakers need to be sensitive to the overall relationship between Goal 5, the measures developed to accomplish it, and the impact that scoring and reporting such measures is likely to have on the various players. "If only a single report were issued, for the nation as a whole, then a single standard might make more sense," predicts Dunbar. "In either case, the system of achievement standards should be *highly sensitive to the nature of the reporting system that NCES develops*."[417] *(Emphasis added)* Validity is a continuing concern, and Dunbar foresees the possibility that the political impact on the postsecondary infrastructure could lead to the worst excesses of "teaching to the test," and thereby compromise the very real gains it was intended to foster.

In White's ideal portfolio world, one wants the widest possible latitude in imposing limiting constraints on the judgement process. In more traditional testing environments, however, multiple standards of test construction and judgement could compromise the value of generalizations and conclusions based on the results. Scriven points out that tests relying overly on the jargon of a field can be difficult to base sound conclusions upon. And while his point was not made in the context of the language of testing as it applies to differing cultural backgrounds, its implications should not be lost: "We want to watch ourselves a bit with this, and make clear that there are roughly 72 words in the english language which are in our common vocabulary, which are terms of logical or critical appraisal, and that's a pretty good, rich vocabulary, and we want to stay with it as much as we can."[418]

---

[415] Dunbar, op. cit., 14.

[416] Ibid., p. 14.

[417] Ibid., p. 14.

For the "report," whatever its form, to be meaningful, its descriptions at some point must become comparisons, even if to an array of standards. White and others are concerned that the "same old values" implicit in testing for years will perpetuate racial, ethnic, and class stereotypes, and thereby overlook the valuable and measurable skills possessed by those whose style of expression—even of thought—may differ. The distinction is important, and subtle. Scriven's 72 words may indeed be vital to a reporting system that is to be of wide, immediate use, but [White would probably caution] don't confuse that language of reporting with the richer and more idiosyncratic language of test construction.

Clarifying the Contributions of College

Another issue is how to tie the results of the assessment to conclusions about the implied focus of Goal 5, the contributions of postsecondary institutions. The circularity of the problem cannot be avoided. As Morante's history of New Jersey revealed, real teachers and administrators are inextricably subject to a system of federal funding, local tenure, resource allocation, and other basic elements of the education environment. Regardless of one's emphasis on the potential *value* of a national assessment of college student learning in informing practice and improving instruction, the undeniable fact is that the national assessment of college student learning will focus a lens on the *success* of instruction in college. While who is to be tested and when in their careers are questions that have been deferred for later study, the standards issue does depend to some extent on how definitively critical thinking abilities and their *change* can be attributed to postsecondary experience. Again, leave aside the debate over whether performance is more a reflection of general abilities (possibly that *bête noir*, "intelligence") there remains an issue about the reported outcome of the assessment: when and how were these abilities (whatever they may be) developed?

---

[418] Michael Scriven in Open Session.

Many at the workshop saw the unequivocal solution to this dilemma as admissions testing, even though such a system would then open up a writhing can of worms about gateway phenomena and the impact on the admissions process. Yet, argues Ratcliff, "what good is an assessment program that fails to inform us in ways useful to the improvement of collegiate education?. . .Without parallel admissions testing, we would be unable to determine if the decline was attributable to secondary or postsecondary education," though such a system would "be able to tell if students in any one year did better or worse than students in preceding years."[419]

But such a conclusion, in isolation, would be simplistic, let alone uninformative about the contributions of college learning. Any standards applied to such a series of outcomes would have to be controlled for by any number of variables.

> We know that in years of economic recession and depression more people, particularly unemployed adults and women returning to the workforce, enroll in college. As the proportion of the population that attends colleges increases, the averages of student scores on ability and achievement tests decline. Thus, as colleges and universities serve a broader array of student abilities, the proportion of academically-talented students is likely to grow smaller. A national assessment program needs to be able to distinguish between a decline in the quality of educational programs and an increase in student participation in higher education.[420]

Also contributing to the difficulty of drawing global conclusions is the wide range of backgrounds possessed by applicants to many colleges—especially to public institutions. From a systemic point of view, Ratcliff is forced to make the point that "not all diversity is good, however. The current divergence in the quality of preparation of high school students severely inhibits colleges and universities' ability to foster higher levels of skill in critical thinking, communications and problem solving." Thus, while it is crucial to distinguish gains attributable to the postsecondary experience, "such a program cannot rely alone on assessment programs based in the secondary schools."[421]

---

[419] Ratcliff, op. cit., 14.

[420] Ibid., 15.

[421] Ibid., 4.

Without a baseline for reference, predicts Ratcliff, any decline over a series of years in the national assessment of college student learning scores would cause a confusing brouhaha. "The journalists' penchant for guilt by association would no doubt lead several reporters to assume that colleges were directly responsible for such declines."[422] This when, as a postulate, the declines could in fact be a function of the qualities already possessed when entering college by those who would be tested four years hence. Such thin and simplistic analyses of the problem do not do justice to the underlying complexities. "While there is widespread concern over the academic preparation of students entering higher education, there is a lack of consensus as to the exact nature and extent of the problem. Similarly, there is disagreement as to the strengths and deficits of contemporary undergraduate education as manifest in the abilities of today's college graduates."[423] And even the establishment of a parallel testing regime in high school would not necessarily solve the ultimate problem posed by Goal 5. Such "baseline information would only be available for those students who we already know the most about,"[424] believes Ratcliff.

As Ratcliff suggests above, diversity will necessarily influence scores, especially scores on instruments about which many have, for years, been posing increasingly difficult questions; questions about basic cultural bias, and about how many current tests may reflect much more accurately the nature of the educational system than the abilities of students it produces. He worries, along with a number of others in the dialogue, that a reflexive imposition at the secondary level of some sort of test designed to provide a baseline for a national assessment of college student learning would only provide essentially redundant information about students "who are most likely to perform well on academic achievement tests, and who are best prepared to succeed in college. We miss gathering information on the very students that we strive to encourage, include and provide access to higher education." He invokes the history of higher education in America once again, in this context, to suggest that "similar to the educational challenges of the turn of the century, we once again face a pressing need to set clear standards for the articulation of precollegiate and higher education."[425]

---

[422] Ibid., 14.

[423] Ibid., 5.

[424] Ibid., 15.

[425] Ibid., 15.

The Diversity of Populations

As mentioned earlier in the context of global competition, other western democracies can only aspire to *improve* their postsecondary rates to near 50%, while "as a country we take pride in noting that nearly half of the high school graduates in the United States go on to college."[426] But when it comes to assessment, the so-called non-traditional segments of this postsecondary population have been chronically under-represented. Asserts Ratcliff, "few demonstrably successful testing programs. . .include representative proportions of part-time students, working adult students, minority students in commuter institutions."[427] The nation's largest secondary test, for example, "the Scholastic Aptitude Test, has been repeatedly attacked for alleged bias against women, African American students, American Indian students. By creating a new test," he warns national assessment of college student learning planners, "we are not likely to produce a measure that is less assailable to these charges."[428]

A number of authors and workshop discussants insisted on the need to take a fresh look at some of the assumptions underlying testing and cultural bias. Their concern was to avoid "a national testing program [that] may dissuade or exclude the very individuals for whom the national objective of increasing critical thinking, communication and problem solving skills may be most needed."[429] Paul and Nosich offered a philosophical prerequisite for framing abilities that engages this dilemma:

> We can respect cultural diversity best by constructing tests in higher order thinking that focus on skills and abilities necessary in all modern cultures. In this way we can legitimately justify assessing it in all cultural groups. Basic critical thinking skills and abilities—because they are based on fundamental elements implicit in the structure of all reasoned thought per se, and because their mastery is essential to higher order thinking in all academic, professional, personal, and public life—are an appropriate foundation for assessment.[430]

---

[426] Ibid., 13.

[427] Ibid., 15.

[428] Ibid., 14.

[429] Ibid., 14.

[430] Paul and Nosich, op. cit., 7.

For Boehm, this apparently admirable goal may be impracticable, and in fact wrong-headed: "I think I understand the intention here, but I also think there is a dilemma. I don't believe test designers can (or should) determine which are to be the "common-core skills, abilities, and traits useful in other cultures; it seems best to let other cultures do that piece of determining."[431] Boehm wants to be clear about the limits of trying to rigorously connect decisions made about test content and structure to imputed or postulated values. "What *can* be done is identify "the skills, abilities, and traits useful," even necessary, for succeeding in *this* 'culture,' whether broadly defined as this 'society' or more narrowly defined as a particular workplace, or an academic discipline." Even with such clarification, controversy is unavoidable, and should be anticipated. Said Boehm:

> I think Paul and Nosich are trying to put the best light on the unavoidably dark side of any national assessment of critical thinking. There have to be standards of some sort; they may not be everyone's standards, predictably they won't be; certainly they won't be every culture's standards. Some won't like that. It's unfortunate, but I don't think we should avoid seeing it for what it is.[432]

In addition to race and ethnic diversity, the question of gender also arose in the national assessment of college student learning discussion. Greenberg urged that attention be paid to "what we have learned about gender differences in learning. We have a body of research that is important, relative to how men and women learn, and there are similarities and differences."[433] This, thought Facione, was an example where Nummedal's call for pedagogical reform was warranted, pointing to the indications which "suggest that the typical college level critical thinking course differentially advantages men over women."[434] An even larger caveat comes with the Alverno experience, rich as it is, because of the exclusion of men from the sample.

---

[431] Boehm on Paul and Nosich.

[432] Ibid.

[433] Elinor M. Greenberg in Open Session.

[434] Facione on Nummedal. "The evidence for a differential impact of standard CT instruction by gender is presented in "Technical Report #3, *Gender, Ethnicity, Major, CT Self-Esteem, and the California Critical Thinking Skills Test*" (ERIC Doc. No.: ED 326 584).

Additional information about the California Critical Thinking Skills Test is available in "Using the CCTST in Research, Evaluation and Assessment," (ERIC Doc. No.: TM 017 349). This publication is also available from the California Academic Press, 217 La Cruz Ave., Millbrae, CA., 94030."

f. What does the debate over Portfolio Assessment reveal about the standards and value issues?

White's contribution to this aspect of the debate was seminal, even though it led to a recommendation—in favor of portfolio rather than instrument assessment—with which few others could concur. But often the objections were based on the same general sorts of philosophical problems the workshop had in trying to deal with the richness of the Alverno model. Put starkly: How in the world would we ever put together something so elaborate and—in the case of portfolios—so unwieldy and subjective? Of course it is the very richness of that subjective detail as a better lens through which to view critical thinking that portfolios supporters stressed.

The question, as with the Alverno model, was not black and white, either/or. There is just too much good sense and actual assessment experience at Alverno to ignore: even if the national assessment of college student learning comes out looking very different from Alverno's system, it will inevitably owe much to the insights developed there. So, too with the portfolio idea. While a consensus developed about how nearly impossible it would be to proffer a portfolio system in answer to the current call for a national assessment of college student learning, the arguments in favor of such a system continually echo and elaborate upon many of the premises dissenting authors and discussants believed crucial.

White picks up the point raised above by Ratcliff about continually testing and retesting the same subgroup of American students. "We should not begin an assessment without considering carefully what kind of results we seek. If we are looking for confirmation that students from secure and privileged families do better on tests than students with insecure and deprived families, we will assuredly produce such results; but they will tell us nothing new and be of no use."[435]

Throughout his paper White confines his analysis to the assessment of writing, but nonetheless addresses issues that a number of others believed were relevant to the overall assessment of critical thinking abilities. "We must be particularly careful to avoid perpetuating class, racial, and social distinctions in the name of writing assessment."[436] The writing process, as described by White, can provide a penetrating lens into critical thinking, but only if a crucial distinction is maintained. As described in Chapter 2, that distinction comes from evaluators, who must discern the process of writing (and the thinking that goes into it) from the naked products of writing, which are all too often doggedly analyzed for mechanical correctness. White concedes the utility of such a viewpoint to the rote learning of rules and grammar in elementary school, but cites a number of new ideas now challenging traditional dogma.

---

[435] White, op. cit., 19.

[436] Ibid., p. 7.

> As with any major shift in thinking, there are many causes of this new emphasis on the writing process: the onset of open enrollment in the City University of New York in 1970, research in socio-linguistics and dialectology demonstrating the value of variant dialects, heightened concern for racial bias in instruction and measurement as part of the drive toward a more racially open society, increasing worries about the place of creative thought in an increasingly passive school curriculum, fuller understanding by writing researchers of how writers actually produce texts, the increasing opportunities for employment of independent thinkers as well-paid rote jobs in industry disappear, and so on.[437]

Based on the twin premises that: certain ethnic groups may think in culturally-derived patterns at odds with implicit assumptions made by traditional instrument constructors; and (also that) they may express themselves in dialects which must be sensitively interpreted in order to appreciate their value, White warns that "the surest way to enforce such undemocratic class distinctions is to assess only (or principally) correctness of writing products, in the traditional way, according to the school dialect. We must assert clearly that correct writing is not the same as good writing, despite pervasive social concerns for mechanical correctness."[438]

The equivalent error in the assessment domain, contends White, would be to "administer a reductive multiple-choice or essay test, yielding simple numbers that can be used only for political purposes."[439] He adduces some fundamental facts to buttress his attack on the instrument model, referring to multiple-choice testing:

---

[437] Ibid., 5.

[438] Ibid., 7.

[439] Ibid., 19.

the validity issue becomes even more serious as evidence accumulates that the supposed correlation of scores on these tests with measures of actual writing is much higher for students who call themselves "white" than for those who are part of racial or ethnic minorities.[440] As long as the validity of the multiple-choice tests of writing is so suspect, none of the other advantages should carry much weight.[441]

Another is the recommendation of the Nation Council on Education Standards and Testing (1991) that "endorsed a national assessment system over a single national test to measure student achievement,"[442] which White believes is the only way to properly take account of "the vast differences in dialect, privilege, geography, culture, and almost everything else that appear in the United States."[443] White concurred with those who favored reporting by institution, but stressed that "any comparison of institutions should take into account the wide disparity of student populations in American higher education. Reports should avoid simple and misleading numerical data, but be rich in examples of different kinds of portfolios from different kinds of institutions."[444] To be useful, he believes, standards must evolve intimately with the system they purport to represent. "If we confuse national achievement standards, which can be set in a useful way, with a national test, which cannot, we make a fundamental error, confusing goals with procedures. Any useful assessment system must recognize regional, or even local, assessment responsibilities."[445]

Focusing on the more practical issues, Facione repeats to White the same general reminder he made to earlier claims that the state of critical thinking research and evaluation was rudimentary:

---

[440] See J. Koenig and K. Mitchell, "An Interim Report on the MCAT Essay Pilot Project in *Journal of Medical Education*, 63(1988): 21-29; and

E. White and L. Thomas, "Racial Minorities and Writing Skills Assessment in The California State University and Colleges." *College* English, (42)1981: 276-283.

[441] White, op. cit., 19.

[442] Ibid., 20.

[443] Ibid., 20.

[444] Ibid., 22.

[445] Ibid., 20.

> In his rush to embrace the portfolio strategy, Dr. White gives neither standardized testing nor essay testing a fair hearing. The field of educational testing and measurement does not have to be reinvented. There is a considerable literature which documents the strengths and weakness of each strategy. There are counter-examples to Dr. White's overstated criticisms and ways to overcome the problems his paper cites.

Many of the current tests effectively address higher order thinking skills, says Facione.[446] In particular,

> the newly published California Critical Thinking Skills Test (CCTST) has demonstrated content validity, construct validity, reliability, and is free from gender bias and ethnic bias. The CBEST (California Basic Educational Skills Test), addresses a similar set of skills but at a more primitive level. The research supporting the CCTST illustrates that Dr. White's claims about not being able in principle to test core college level critical thinking skills using objective, standardized testing are simply wrong.[447]

Facione concedes that "experiments in college departments with portfolio assessment of their own students certainly can be made to work. But they do not make the case that the portfolio strategy is the national assessment panacea."[448] Partly because portfolios work best, continues Facione,

> when the persons being assessed come from a common background (e.g., all were once music majors) or from the same program (e.g. all are currently enrolled in an

---

[446] Facione on White: "ETS (Princeton, NJ) can provide basic information about the LSAT, GRE, and SAT. The American College Testing Service can provide information about the Advanced Placement program in composition. Other national testing services and regional educational research labs are good sources of additional information. The thousands of publications on student assessment, program evaluation, and educational testing are so vast and varied that there is an entire ERIC Clearinghouse devoted exclusively to this concern."

[447] Ibid. "The four Technical Reports supporting validity and reliability of the California Critical Thinking Skills Test, published in 1990, are ERIC documents: ED 327 548 "Experimental Validation and Content Validity," ED 327 550 "Factors Predictive of Critical Thinking Skills," ED 326 584 "Gender, Ethnicity, Major, CT Self-Esteem, and the California Critical Thinking Skills Test, and ED 327 566 "Interpreting the CCTST, Group Norms, and Sub-Scores".

[448] Ibid.

> international studies capstone seminar course) and when all the judges who jury the portfolios have well developed and commonly accepted standards (e.g. all have expertise in the area to be assessed, all agree to use the same grading rubric and criteria, and all been trained and have practiced on sample cases to refine their inter-rater reliability coefficient.)[449]

Since these two conditions cannot be met on a national scale, Facione concludes that adequate validity and reliability are a pipe dream for a national portfolio system. If the judgement cannot be systematized, the public ultimately—and correctly, insists Facione—will come to perceive the process as unfair. Ironically, though White seems to foresee a more complex system of reporting with many more subjective variables and frames of reference, the portfolio system might in fact be subject to an enormous bias.

There does, however, seem to be a middle ground, which Facione concedes in his objections to White's portfolio idea: "essay testing at least starts from a common point of reference for all those taking the test—the well conceived question prompt."[450] Chaffee agreed, saying that while "portfolio assessment is an emerging methodology," its complexity "raises serious questions regarding its feasibility for [a national assessment of college student learning]. Perhaps some of the key elements of this approach might be integrated with a more traditional essay test—for example, giving students the opportunity to revise their essay in response to new information or some form of feedback."[451] But even here, validity is hard to achieve. As Larson explains, in his commentary on the New Jersey experience,

> The validity problem [arises in] the contamination of the scores by essay-writing performance. While expression is rightly seen as one of the key higher-order skills, the facile writer tends to garner too many points for it at the expense of the logically crucial matters, as we confirmed in the Carnegie evaluation of the National Writing Project.[452]

---

[449] Ibid.

[450] Ibid.

[451] Chaffee on White.

[452] Larson on Morante.

Larson suggests the use of so-called multiple rating items to avoid "getting out of the frying pan of multiple choice items [only to] get into the fire of essays."[453] He explains how this type of item can "bypass the usual flaws in multiple choice items but retain fast scoring properties:

> A simple example of a multiple rating item requires the testee to allocate a grade to each of a set of answers to a common question; any of the answers can in principle get any grade so the 'lesser evil' algorithm doesn't work. More complex examples provide a repertoire of several critical or positive verbal comments, so that a more elaborate response can be constructed by checking the letters for more than one response; e.g., a grade and a reason for it. In either case, we are of course dealing with higher level skills, evaluation, synthesis, etc., and not recognition skills. Of course the grading scale is defined, some of the items provide possible reasons for the grade, occasional write-in can be allowed. One can also allocate half marks for grades that are adjacent to the correct grade and alter the marking rubric so as to penalize answers which show a total lack of understanding, etc.[454]

Such detailed analysis of test item types begins to bleed into the major area of national assessment of college student learning test construction, which has been postponed for consideration in a subsequent paper. But the general drift towards portfolios as a panacea to the many difficult issues involved in national assessment of college student learning raised—in those familiar with the research—a compulsion to warn. Cappelli's workplace analysis, noted Miller, led him to conclude that "work samples are the 'selection device closest to achievement tests,' and [he] maintains that one could make use of existing student work (like research papers, laboratory experiments, and the like) to assess college graduates' performance."[455] Several very fundamental concerns about such approaches recur:

- Averaging the grades on written work, as Dr. Cappelli suggests, takes us right back to the problem of the lack of correlation between grading and job performance. Portfolios need to be assessed according to a scoring guide that mirrors the skills, knowledge and abilities we want to measure. They are not themselves an assessment measure but merely the raw material to which the measure—the scoring guide—is applied. That implies the training of assessors.

---

[453] Ibid.

[454] Ibid.

[455] Miller on Cappelli.

- Portfolios, like work samples, capture only a small range of skills, though the range might be enough to include writing, critical thinking, and problem solving. This is one reason why The College of William and Mary, which Dr. Cappelli mentions as an example of an institution that has used this approach to the assessment of general education, has in fact abandoned it.
- I am not aware of tests of validity or reliability that have been done on portfolio assessment.
- Finally, they are a logistical nightmare to collect for large programs, even using sampling procedures, which is why they have turned out to be more useful for assessment in the major than assessment of general education. It would be very hard to collect comparable portfolios across higher education in the country, and the process would be expensive. And how do we factor into the overall report on student performance the lack of certain kinds of work in the students' portfolios? The College of William and Mary, for instance, found that they had trouble getting portfolios with enough written work in them to do even a crude assessment of writing skills."[456]

In the end, this discussion will not produce a clear-cut answer. As Boehm noted in praising Paul and Nosich's catalog of affective dimensions, "(intellectual perseverance, for example), any testing would have to take place over an appropriately long period of time and thus [they] could not be legitimately assessed at all during the time-frame suitable for a national test."[457] "Perhaps," said Boehm, but

> without, for now, addressing the value, or the lack of value, of assessing 'thinking independently,' or 'intellectual perseverance' or 'intellectual courage,' or any of the others, it seems pretty clear to me that these *can* be assessed very effectively by portfolio, which allows for, even calls for, a variety of assessment materials,

---

[456] Ibid.

[457] Paul and Nosich, op. cit., 16.

including, especially, drafts of essays on a range of topics, which reflect the student's thinking process as well as his disposition toward thinking.[458]

* * *

On balance the question doesn't seem to be whether portfolios provide insight into thinking that is not as easily—or at all—gained through another method. Rather, national assessment of college student learning developers must struggle with how (assuming their potential value) to defend and properly report such insights, even if the resources could be found to establish a system to collect and develop them. Thus the issue comes full circle, as framed in Chapter 1. What kind of national assessment of college student learning can we do *now*, with the ever present limits of time, money, and political will? Or, as some say, must we re-frame the question, and resist the temptation to provide a system to politicians and educators who are willing to fund a national assessment, but only one that they can defend by fairly conservative criteria. Should we hold out for a richer and deeper analysis of Goal 5 and critical thinking, and try to establish a revolutionary system of research, development and implementation, predicated on the feedback loop to inform instruction throughout the American educational system.

---

[458] Boehm on Paul and Nosich.

4. SUMMING UP

Given these considerations, can the national assessment of college student learning proceed?

This riot of premises and possibilities raised in Chapter 1—some conflicting, others confluent—presents national assessment of college student learning developers with the clear-cut challenge of erecting a frame and a foundation that will withstand considerable criticism. ***Robert Ennis*** of the Illinois Critical Thinking Project at the University of Illinois is regarded by many as the "father" of this movement in modern America, and indeed his pedigree is longstanding. "In 1958, my doctoral dissertation was entitled "The Development of a Critical Thinking Test." I've been working on the problem ever since, and am delighted to see the great interest that has developed in this area."[459]

Ennis participated in the NCES workshop, and wanted "to try to resolve this basic issue of the purpose of this operation as it relates to Goal 5: Is the purpose Assessment, or is it Improvement of Instruction, as the issue is put. . .I want to suggest that it can be both, but to do so in a way such that some of the problems that would develop if the same administration were used for both purposes might be avoided."[460] He began with the assumption that there will be a "very high quality set of devices or instruments" developed to answer the fundamental question—Has there been improvement in these three areas, and how much?" [Although the structure of the instrument and the issue of motivation are not included in this report, it should be noted they were widely debated at the workshop, and to some extent "bleed over" into this summary discussion.]

Seeking an answer, Ennis crystallized the deeply felt concerns of many by acknowledging that if *Improving Instruction* is the driving force, the national assessment of college student learning will of necessity have to be administered to "many, many more students" than if *Accountability* were the master. Among the many issues, thought Ennis, raised by the specter of an *Accountability* model, two of the most serious seem to be the cost/complexity of administration and the inevitable "high stakes" atmosphere that will arise.

---

[459] Robert Ennis in Open Session.

[460] Ibid.

In response to Ennis' concern with their limitations, Michael Scriven said of going beyond matrix sampling:

> It doesn't [have to] lead you into the problems that Bob was warning us about. In matrix sampling, as he rightly pointed out, you can't give a very enlightening feedback to the individual. But if you go for full tests for each of your sample, but do not undertake to take a large enough sample from institutions—or for that matter from states—so that you can give a report at the institutional or state level, you don't get into high stakes.
>
> What you do get is really important: you get an [institutional] incentive to participate. Which is something we've got to take extremely seriously from Day One. And the incentive is 'Here are some really important skills. If you can give us a certain amount of your time, we'll give you feedback on your performance on those skills, and we'll give you a certified transcript which you may—at your option—use in applying for jobs.' Now there's no need at all for that to be treated as high stakes for institutions, for instructors, etc. But we should, I think, make the tests available in some format—a parallel form of them—so that instructors, and for that matter institutions that wish to participate in having some institutional measurement made, can do so, too.[461]

Ennis' conclusion may itself be a paradigm of critical thinking, for he refuses to accept the problem and perhaps unavoidable paradox as stated, but instead reframes and redefines the goal:

> Now I think there are four ways, at least, that we can [conceive] an instrument or set of devices that is primarily a monitoring kind of instrument. I also want to distinguish between monitoring and accountability. Monitoring is really what the goal calls for. It doesn't call for us attributing responsibility to the states or to institutions, which is what accountability would do. If we just know what the level is, then there's much less chance of high stakes.
>
> So, instead of just having two choices, accountability or educational improvement, I think we really have three choices: monitoring, accountability, and instructional improvement. What I would like to urge is that we use the instruments for

---

[461] Michael Scriven in Open Session.

> monitoring as well as instructional improvement, and want to suggest four ways in which the monitoring instrument could be used for instructional improvement. Not directly, but indirectly.[462]

First, citing how *Nation At Risk* took the results of NAEP "and used them to advertise the alleged deficiencies in the reasoning—among other things—of the students," Ennis notes that such a message can have "a tremendous impact." Second, the mere announcement of the goals embodied in a national assessment of college student learning provides a model for individual institutions. Third, echoing the hope expressed by Greenberg, Ennis envisions many in the educational community saying of a national assessment of college student learning "'Hey, that's a good idea. Let's try something like that locally.' And they could use it for local accountability, and they could use it for local experiments, research, and local feedback."[463] Finally, an idea that picks up on the research agenda of many who couldn't see how to proceed now, "this monitoring instrument could be used in small research studies, which then would have feedback into our techniques for teaching at the higher level."[464]

Ted Marchese seconded Ennis' suggestion that "this effort ought to put critical thinking, problem solving, and communication abilities into the public and the institutional and faculty minds in a way that it is not now.

> Now that's the very important kind of thing. If it makes faculties, and institutions, collectively aware that these are important things that should happen in undergraduate education, and raises demand within them for ways of arranging curricula and pedagogy so that they're more likely to occur, that's the larger outcome, rather than particular feedback to me as a teacher of sophomore organic chemistry, on how to fix my course. It's not that. The feedback that we want is that we want to put these three things more firmly in the public mind, and in the faculty mind, and have programs come behind that tell people how they can more likely achieve these kinds of outcomes.[465]

---

[462] Ibid.

[463] Ibid.

[464] Ibid.

[465] Ted Marchese in Open Session.

Marchese continued to remind the workshop to consider how this creation they were contemplating would appear to so many different audiences, and how pragmatic they needed to be in considering not only how their conception would look, but how the actual national assessment of college student learning—fully blown but perhaps politically modified from the game plan—would work. Ewell applauded this pragmatism, and cautioned his fellow scholars about an assumption that was too often going unexamined during the discussion, the assumption that

> we're going to have infinite time and money to do this. I think that the major difficulty that you can see in the experience of states in [their] efforts of assessment in higher education, [is] that you have to look at whatever program you design, as though it were half implemented. You have to look at it as though the wonderful thing you've put together is going to have to be implemented with half the resources that you expected, in half the time you expected, with a great deal of political interference, with a great deal of special interest lobbying, with a great deal of modifications to that design, that were not taken into account at the beginning.
>
> I think that we've got to be designing a set of instruments, approaches, and devices which are very robust. They are things which will give us some information in spite of the difficulties of half implementation and political interference. That's not an easy task. But it's definitely one that we should be paying a good deal of attention to.[466]

As is often the case in effective critical thinking, the value of the answers depends on the precision of the question. There was no dearth of questions during the workshop, and questions about questions. Since the process was defined at the outset, as inclusive, scholars were encouraged to cast a wide net lest something important be overlooked. While Greenberg didn't lay the issue to rest, she focused on some of the more irreducible of these—"Where do states fit? Moving from the learner out of the institutions. . .into the societal question. Who are the stakeholders? Why was the question asked? and Who is the client, or clients, as the case may be.

> This is a real opportunity for the role of the department, and NCES in particular, to shift, and to become a model for collaboration, in a way in that institutions don't typically do. When you have the lever, and you can't control the nodes in the

---

[466] Peter Ewell in Open Session.

> system—which you can't in this instance—then one must figure out how to manage the network organization in a very different way.[467]

In closing it should be noted that if there was a consensus, it was that although assessing college student learning can be complex and is fraught with numerous potential pitfalls, it can be done. Further, based upon the concerns and suggestions of workshop participants there needs to be some notion of the expectations of college student learning in light of Goal 5. Operationally this suggests that a logical starting point is the identification and justification of the potential skills and attributes to be assessed and related levels of proficiency. [468]

---

[467] Elinor M. Greenberg in Open Session.

[468] Based upon these suggestions, identification of a potential set of skills, along with related levels of proficiency that might be used to assess a students progress, is the focus of the second study design workshop.

APPENDIX

LIST OF PARTICIPANTS

Trudy Banta, University of Tennessee, Knoxville, Tennessee
Paul Barton, Educational Testing Service, Princeton, New Jersey
Nancy Beck, Educational Testing Service, Princeton, New Jersey
Ernest Benjamin, American Association of University Professors, Washington, D.C.
Lorenz Boehm, Oakton Community College, Des Plaines, Illinois
Cornia Burmer, Center for Children and Technology, New York, New York
Robert Calffe, Stanford University, Stanford, California
Wayne Camara, American Psychological Association, Washington, D.C.
Magda Colberg, Office of Personnel Management, Washington, D.C.
Delinda Cannon, Midland Technical College, West Columbia, South Carolina
Peter Capelli, The Wharton School, Phildelphia, Pennsylvania
John Chaffee, LaGuardia Community College, Long Island City, New York
Pat Courts, SUNY Fredonia, Fredonia, New York
Allen Doolittle, ACT, Iowa City, Iowa
Steve Dunbar, University of Iowa, Iowa City, Iowa
Robert Ennis, University of Illinois, Champaign, Illinois
T. Dary Erwin, James Madison University, Harrisonburg, Virginia
Peter Ewell, National Center for Higher Education Management Systems
Peter Facione, Santa Clara University, Santa Clara, California
Norman Fredericksen, Educational Testing Service, Princeton, New Jersey
Ed Fuentes, National Education Goals Panel, Washington, D.C.
Anthony Golden, Austin Peay State University, Clarksville, Tennessee
Elinor Miller Greenberg, EMG Associates, Littleton, Colorado
Diane Halpin, California State University, San Bernadino, California
Ron Hambleton, University of Massachusetts, Amherst, Massachusetts
Joan Herman, UCLA Center for Evaluation, Los Angeles, California
Buddy Karelis, FIPSE, Washington, D.C.
Mike Knight, Kean College, Union, New Jersey
Richard Larson, Lehman College, Bronx, New York
Donald Lazere, California Polytechnic, San Luis Obispo, California
Charles Lenth, State Higher Education Officers Organization, Denver, Colorado
Georgine Loacker, Alverno College, Milwaukee, Wisconsin
Irene Lynn, Department of Labor, Washington, D.C.
Ted Marchese, American Association of Higher Education, Washington, D.C.

Jean McDonald, National Governors Association, Washington, D.C.
Aims McGinnis, State Higher Education Officers Organization, Denver, Colorado
Marcia Mentkowski, Alverno College, Milwaukee, Wisconsin
Margaret Miller, Virginia State Council of Higher Education, Richmond, Virginia
Ed Morante, College of the Desert, Palm Desert, California
Gerald Nosich, Sonoma State University, Rohnert Park, California
Susan Nummedal, California State University, Long Beach, California
Richard Paul, Sonoma State University, Rohnert Park, California
Natalie Peterson, University of Pittsburgh, Pittsburgh, Pennsylvania
James Ratcliff, Pennsylvania State University, University Park, Pennsylvania
Daniel Resnick, University of Pittsburgh, Pittsburgh, Pennsylvania
Donald Rock, Educational Testing Service, Princeton, New Jersey
Rebecca Rubin, Kent State University, Kent, Ohio
Michael Scriven, Pacific Graduate School of Psychology, Palo Alto, California
Larry Suter, National Science Foundation, Washington, D.C.
Ron Swanson, Texas Academic Skills Program, Austin, Texas
Mary Tenopyr, AT&T, Morristown, New Jersey
Richard Venezky, University of Delaware, Newark, Deleware
Norman Webb, University of Wisconsin, Madison, Wisconsin
Mark Weinstein, Montclair State University, Upper Montclair, New Jersey
Edward White, California State University, San Bernadino, California
Joan Wills, Center for Workforce Development, Washington, D.C.
John Wirt, SCANS, Washington, D.C.
Barbara Wright, American Association of Higher Education, Washington, D.C.

☆ U.S. GOVERNMENT PRINTING OFFICE:1993-342-109/83121

United States
Department of Education
Washington, D.C. 20208–5653

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G-17

FOURTH CLASS BOOK RATE

BEST COPY AVAILABLE

192